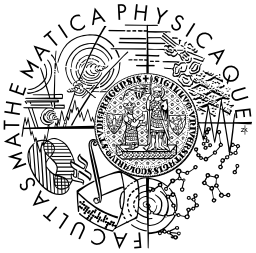
Charles University

Faculty of Mathematics and Physics

DOCTORAL THESIS



FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

Mgr. Filip Děchtěrenko

Comparison of scan patterns in dynamic tasks

Department of Software and Computer Science Education

Supervisor of the doctoral thesis: Mgr. Jiří Lukavský, Ph.D.

Study programme: Computer Science

Study branch: 4I1

Prague 2017

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In Prague 28.4.2017 signature of the author

Title: Comparison of scan patterns in dynamic tasks

Author: Mgr. Filip Děchtěrenko

Department: Department of Software and Computer Science Education

Supervisor: Mgr. Jiří Lukavský, Ph.D., Department of Software and Computer Science Education

Abstract: Eye tracking is commonly used in many scientific fields (experimental psychology, neuroscience, behavioral economics, etc.) and can provide us with rigorous data about current allocation of attention. Due to the complexity of data processing and missing methodology, experimental designs are often limited to static stimuli; eye tracking data is analyzed only with respect to basic types of eye movements – fixation and saccades. In dynamic tasks (e.g. with dynamic stimuli, such as showing movies or Multiple Object Tracking task), another type of eye movement is commonly present: smooth pursuit. Importantly, eye tracking data from dynamic tasks is often represented as raw data samples. It requires a different approach to analyze the data, and there are a lot of methodological gaps in analytical tools.

This thesis is divided into three parts. In the first part, we gave an overview of current methods for analyzing scan patterns, followed by four simulations, in which we systematically distort scan patterns and measure the similarity using several commonly used metrics. In the second part, we presented the current approaches to statistical testing of differences between groups of scan patterns. We present two novel strategies for analyzing statistically significant differences between groups of scan patterns and show their application in two behavioral experiments. In addition, we also showed an example of a classical approach to testing differences between groups of scan patterns to answer the question about eye data quality. In the final part of the dissertation, we predicted scan patterns in Multiple Object Tracking task using neural networks. Our results outperformed current, state-of-the-art methods.

Keywords: eye movements, Multiple Object Tracking, group comparison, scan patterns, neural networks

This thesis is dedicated to my parents, for their support over the years and to my beloved Petra who gives my live a purpose.

I would like to thank to my supervisor Jiri Lukavsky, who taught me a lot during my studies. He was always there whenever I needed a support. I would also like to thank to Cyril Brom for his various advices and for the opportunity to do interdisciplinary research. Finally, I would also like to thank all my friends which supported me in last few years.

Contents

Introduction

The world around us is a complex environment. Every living organism, in order to function properly, needs to process information and modify its behavior based on perceived information. For visual modality, eyes are the organs for perception. Humans usually fix their eyes on the location that they are currently processing. With the development of new technologies, it has become possible to measure eye movements and quantify where participants are looking during various tasks. In the last 20 years, eye tracking has become part of experiments in many different fields: psychology, neuroscience and behavioral economy. Today, eye trackers are able to sample the position of an eye gaze with high frequencies (1000 Hz), which allows researchers to measure subtle changes in spatial position in time. This spatio-temporal representation of eye data is called a scan pattern. The property of scan patterns has been studied in the past (Yarbus, 1967; Noton & Stark, 1971; Brandt & Stark, 1997; Foulsham & Kingstone, 2013).

Traditional studies that represented eye data as scan patterns used static stimuli. For example, the task could be to find one particular item in a display of multiple items. Use of static stimuli has one advantage: when perceiving static stimuli, there are only two main events in eye movements – saccades (rapid eye movements) and fixations (intervals between saccades in which the eyes are relatively still). There are algorithms for extraction of such events from the raw eye tracking measurements (Andersson, Larsson, Holmqvist, Stridh, & Nyström, n.d.) and there are a lot of approaches to analyzing scan patterns that are formed as a succession of events (Le Meur & Baccino, 2013). Although static stimuli can be used in many important studies about human perception, the world is a dynamic place. Our everyday experience is continuous as are the perceived stimuli. One particular example of dynamic stimuli that can be used in experiments is a movie clip. There is a lot of semantic and visual content in each clip and therefore studying eye movements while showing movies would be a interesting source of information (Loschky, Larson, Magliano, & Smith, 2015; Dorr, Martinetz, Gegenfurtner, & Barth, 2010). There are two main problems with using dynamic stimuli in experiments. The first is the presence of another type of eye movement: smooth

pursuit (Holmqvist et al., 2011). This type of eye movement is executed when the task is to track a moving object. Detection of smooth pursuit eye movement is a challenging task, because there is a lot of occulomotor variability that makes the distinction between fixation, saccades and smooth pursuit difficult (Komogortsev & Karpov, 2013). A second, and more important, problem is a lack of methods that could compare scan patterns with smooth pursuit eye movements. Traditional measures used in the eye tracking context either collapse the scan pattern across the time dimension (e.g., Mannan's metric), which results in losing correspondence between a particular time sample and the content of the scene, or they have difficulties in using scan patterns that contain smooth pursuit (e.g., ScanMatch).

As smooth pursuit is a common type of eye movement in dynamic tasks, scan patterns are mostly continuous with introduced discontinuities by saccades. Therefore, we could apply comparison methods outside the eye tracking context. For example, we could conceptualize the scan patterns as time series and compare them using General Lineral Modelling common in fMRI research (Monti, 2011) or by Dynamic Time Warping (Berndt & Clifford, 1994). To increase the use of dynamic stimuli in experiments, we need to study the properties of methods that compute scan pattern similarity across the time dimension as well. There are several metrics that are used for comparison of scan patterns from dynamic tasks, but there are some open questions regarding them. How do the metrics scale when scan patterns differ in their variability? How is scan pattern similarity evaluated by one metric related to the similarity evaluated by another metric? To study such properties, movie clips are not ideal stimuli. For each frame of the movie clip, it is hard to relate image content to the fixation location; the content of each video frame is perceptually and semantically complex. In addition, when studying the variability of scan patterns, we could only compare scan patterns between subjects. When the same stimuli are presented to the participant, his behavior would change, because he would recognize the movie and might decide to look at different locations during the time course in order to see additional content in the movie. There is, however, one dynamic task that does not have this problem. This task is called Multiple Object Tracking.

In Multiple Object Tracking (MOT) task, participants tracks several moving indistinguishable objects (Pylyshyn & Storm, 1988). This task can be imagined as watching several footballers moving across a field. Several of the moving objects are targets; the others serve as distractors. When eye movements are measured in this task, participants usually look somewhere between the tracked objects. Therefore, we could relate the position of the objects to eye gaze. By varying object trajectories, we could alter scan pattern similarity. The main advantage of MOT is the possibility to present identical trials repeatedly without noticing. This allows us to measure the within-subject variability of a scan pattern. Therefore, we could use it to study the behavior of metrics for scan pattern similarity. Another great advantage of MOT is the possibility to apply geometrical transformations to object trajectories. This could not be applied to the movie in the same way. For example, we could present the trial backwards in time and it would look natural, while playing movies backwards would be noticed for events such as walking backwards. Such properties make MOT an interesting task for studying the properties of time series. As the scan patterns in MOT are influenced by both top-down (the task has a given goal) and bottom-up (eye movements are influenced by the position of the objects), we could measure the scan pattern variability as dependent on noise (either random or systematic).

In this thesis, we focused on three research questions. In the first part, we studied the properties of several metrics used for comparing scan patterns from dynamic tasks. In a series of 4 simulations, we showed the properties of 5 metrics on artificial scan patterns in various experimental scenarios. As the scan patterns resemble time series, we applied one of the methods that is typically used in a different context for scan pattern comparison. We present how the metrics evaluate the similarity of two scan patterns when we systematically distort them. This provides us with guidelines on how one could interpret the coherence of groups of scan patterns.

In the second part (and in fact, this is our main contribution from this thesis), we studied one open problem: how to statistically test differences in scan pattern similarity between groups. This is a tough question; we needed to test whether the variability *between* groups is different from the variability *within* each group.

So far, there is only one method from Feusner and Lukoff (2008) that addresses this problem. The main contribution of this thesis is the creation of two methods for testing similarity between groups of scan patterns. We created simulated data and evaluated the performance of those methods on them. To show the use of these method in a real experiment, we applied one of the methods to test whether we could present flipped trajectories in MOT tasks to mask the repetition. In addition, we used MOT for a question about eye data quality; in particular, whether we can safely ask participants with myopia to remove their glasses without any loss of quality for eye data. We note that our method for the statistical testing of differences between groups of scan patterns can be generalized for other time series.

In the final part, we were inspired by the traditional use of MOT for studying eye movements. When participants track targets in MOT, they look somewhere between the objects. Past research shows many approaches to the modeling of eye gaze in this task (Fehd, 2009; Lukavský, 2013). Here, we tried to model eye gaze in MOT using a data-driven approach. We trained several feed-forward, neural networks on datasets from three behavioral MOT experiments. We used different feature vectors and several methods for how to reduce variability in the dataset.

Organization of the thesis

This thesis is divided into four chapters. The first chapter contains information about the visual systém and eye tracking measurements. It presents the MOT task including the models behind the tracking and an overview of current eye movement models in MOT. Due to the multidisciplinary nature of this thesis, we decided to describe the parts regarding eye tracking in greater detail to explain better the concept of eye tracking: even for researchers outside the vision science community.

The second chapter focuses on the comparison of scan patterns. First, we present the overview of methods that can be used for comparing scan patterns. Then, we describe our simulations in which we present the Correlation Distance Metric

(which we will use in the third chapter) and show its behavior in relationship to different metrics.

The third chapter, the main one in this thesis, focuses on the statistical testing of differences between groups of scan patterns. In three behavioral experiments and one simulation experiment, we introduce two extensions of a current method for testing the significance between groups, and we show its application in real-life scenarios.

In the final chapter, we focus on the prediction of eye movements in MOT using feed-forward neural networks.

Notations used in the text

We are following guidelines from APA 6$^{\text{th}}$ edition (`http://www.apastyle.org/`). Therefore, we usually round to two decimal places with the exception of p values, which are rounded to three decimal places. All values that, in absolute terms, cannot exceed one are written without the leading zero. We report effect size only in Experiment 1, which employs a traditional analysis. In other experiments, we did not use conventional effect size measures, because there is no consensus on which effect size measure we should use for the analysis presented in the text.

1. Visual perception and eye movements

During visual perception, humans process information by allocating their attention across a perceived environment. Although attention and eye movement can be dissociated (Posner, Snyder, & Davidson, 1980), people usually look at the places which interest them. Therefore, studying eye movement can help us answer the question, how do people allocate their attention with respect to the content of a given scene. Eye tracking is a common part of many experimental designs: both for static and dynamic tasks. Static tasks present one static display in each experimental trial (for example, searching for a letter H among a bunch of Ts), while with dynamic tasks the content of the display changes (for example, searching for a dead fly among moving flies). One dynamic task in particular could serve as a paradigm suitable to methodological studies concerning eye movement. This task is called Multiple Object Tracking. Findings on eye movement from this laboratory task could help us deepen our knowledge of complex natural tasks, such as watching video clips.

1.1 Chapter description

In this introductory chapter, we present eye as an organ for sight, including its anatomy and physiology and different types of eye movement. Then we describe eye tracking in general and show how eye tracking is typically conducted using one particular type of eye tracker. In the final part of the chapter, we describe in detail the main task that we will use for various experiments throughout the thesis.

1.2 Visual angle

In vision research, stimuli sizes are denoted in degrees of visual angle (DVA or °). Classical sizes such as pixels or cm would be dependent on viewing distance,

because the same stimuli perceived from different distances would result in a different experience. Visual angle stays the same even when different viewing distances are used. Degrees of visual angle can be computed from viewing distance and pixel size using the formula:

$$\theta = 2 \arctan(\frac{0.5x}{d})$$

where x is stimuli size in cm, d is the distance of the eye from the center of the stimuli and θ is stimuli size in degrees of visual angle (See Figure 1.1). Usually, the sizes are given in pixels and therefore pixels need to be converted first to cm. We will be using the symbol °throughout this thesis: except for in Chapter 2, where we will use both DVA and °.

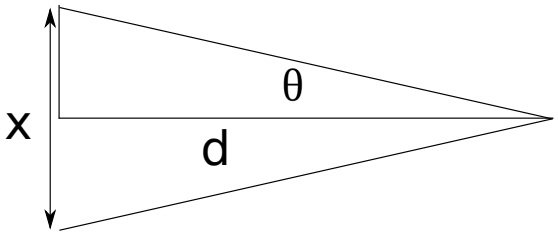


Figure 1.1: Conversion from distance unit (cm) to degrees of visual angle. Details are in the text.

1.3 Processing of visual information

In the section, we present how visual information is processed in the human system of sight. We describe the eye's gross anatomy and the further processing of visual signals. This chapter was developed based on basic information about eye taken from (Bear, Connors, & Paradiso, 2007).

1.3.1 Anatomy of the eye

The eye is an important organ for visual perception (see Figure 1.2). Light enters the eye and goes through the central part called pupil. The size of the pupil is controlled by the iris. Both the iris and the pupil are covered by a transparent surface called the cornea. The area between the cornea and the pupil is filled with an aqueous humor to humidify the cornea. On the inner side of the pupil,

there is another transparent surface (lens), which is controlled by the zonula fibers that change its refractory properties. The inside of the eye is filled with the vitreous humor that transfers light rays to the retina: the inner layer around the sphere that makes up the eye. The outer part of the eye is called the sclera, and it forms the eye's outer barrier. There are three antagonist pairs of muscles attached to the sclera; these guide the eye. On the eye's surface, there are blood vessels that nourish the eye. They originate from an area called the optic disk, where the eye is connected to the brain via the optic nerve. There is an area with almost no blood vessels that processes the central part of the visual field called the macula. In the central part of the macula, there is small dip called the fovea. On the retina, there are specialized cells organized in layers. The specialized

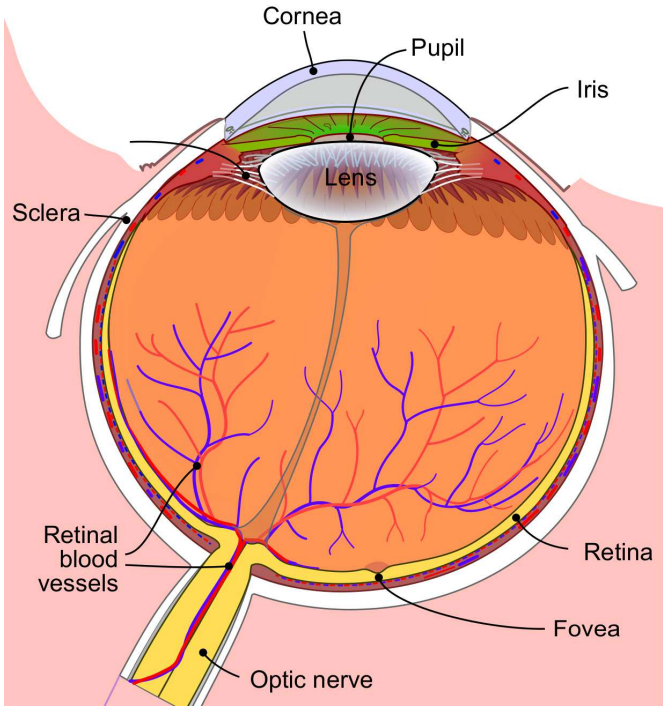


Figure 1.2: Schema of the human eye. This picture was taken from `https://commons.wikimedia.org/wiki/File:Schematic_diagram_of_the_human_eye_en.svg` and adjusted for the purposes of this work

cells for detecting light waves (photoreceptors) are located on the bottom layer of the retina. They transform light into an electric signal. This signal is collected via the ganglion cells located in the top layer of the retina. The ganglion cells transfer the signal further to the brain. There are two types of photoreceptors that differ in their properties. First, there are rods, which are sensitive to light

under scotopic conditions (nighttime vision) and cones, which are much more sensitive to light during the daytime. The rods and cones are distributed non-uniformly across the retina. As visualized in Figure 1.3, there are no rods in the fovea; whereas, the density of the cones is the highest in the fovea. On the other hand, the rods are found mainly in the peripheral retina, where there are almost no cones. There are no rods and cones on the optic disk, so this area is also called the blind spot. There are three types of the cones specialized in reading three different light wavelengths: blue ($\lambda = 430$ nm), green ($\lambda = 530$ nm) and red ($\lambda = 560$ nm).

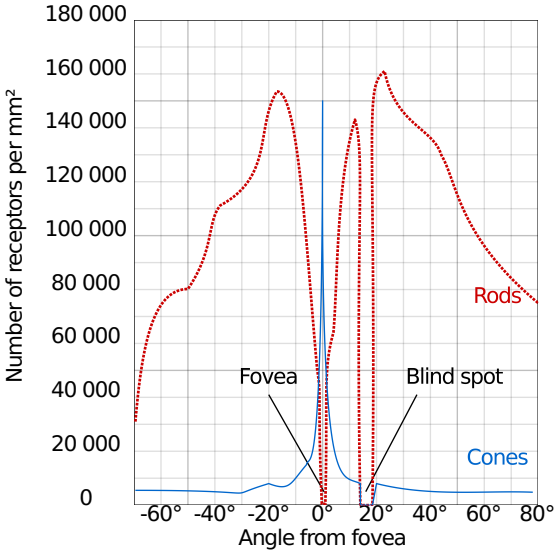


Figure 1.3: Distribution of the rods and cones. This picture was taken from `https://commons.wikimedia.org/wiki/File:Human_photoreceptor_distribution.svg`

1.3.2 Processing of visual information

As reflected light from an object goes through the lens, it projects onto the retina (the original image is turned upside down due to the laws of optics), where the photoreceptors transform light waves into membrane potential. This signal is transferred via bipolar cells to ganglion cells which are the only cells that generate action potential. The signal from neighboring photoreceptors is also integrated by horizontal cells, which transfer the signal to the bipolar cells as well. Each bipolar cell has its receptive field from which responses from the multiple pho-

toreceptors are integrated. A typical receptive field is circular with a central part that is either sensitive to the light (ON cells) or to the dark (OFF cells) which are connected to the photorecptors directly. The peripheral part is connected to photoreceptors via integrated information from the horizontal cells. The central and peripheral part of bipolar cells' receptive fields are antagonists. Therefore, in the case of an ON cell, when illuminated, the central part of the membrane is depolarized while the peripheral part is hyperpolarized. This allows the visual system to form a response on the edges as they fall into multiple receptive fields. The signal from each eye is sent through its optic nerve and the optic nerves cross at the optic chiasm which is located at the bottom of the brain below the hypothalamus. As each eye contains visual input from both visual hemifields, the chiasm serves as a connection point, where the axons transferring the information from the same visual hemifield continue together. After the chiasm, the corresponding parts of the visual field continue together to the lateral geniculate nucleus (a layered structure of neurons) and further on to the V1, which is the primary visual cortex located in the occipital lobe. The fovea plays a crucial role in visual perception. Although it spans less than 2°, visual information is prioritized in the fovea (Hubel & Wiesel, 1974). Due to the cortical magnification (more neurons represent the angle of the visual field), about 25% of the visual cortex processes information from the central part of 2.5° (De Valois & De Valois, 1980). This makes fovea important for the visual perception.
This makes the fovea important for visual perception. A more detailed description of visual information processing is out of the scope of this thesis.

1.3.3 Types of eye movements

During perception, the eye is oriented so that light from the attended location falls onto the area of highest acuity (the fovea). The eye is oriented using the muscles attached to it. Three antagonist pairs of muscles control horizontal (yaw), vertical (pitch) and torsional (roll) movements (Bear et al., 2007). Eye movements can be classified into several types. The main type of movement which moves the fovea from one position to another is called *saccade*. Saccades are fast movements (velocity ranges $30 - 500°/s$) lasting 30–80 ms. Saccades can be classified into

several types based on intended goal (Rommelse, Van der Stigchel, & Sergeant, 2008) or by saccade latency (Fischer & Ramsperger, 1984). Due to biological limitation, the trajectory of saccades between two points is not a straight line. It usually follows several types of curvatures based on the location of the saccade target or saccade type (Smit & Van Gisbergen, 1990). It is also modulated by an instruction to attend to a location other than to the target (Sheliga, Riggio, & Rizzolatti, 1994). The saccade usually does not land directly at the desired location, so it is usually followed by small corrective eye movements called the *glissades* (Holmqvist et al., 2011). Saccades are almost exclusive during static tasks.

The intervals between saccades are called *fixations*. During fixations, the eyes are relatively still, while information from the fixated region is being processed by the visual system. There are three types of eye movements during fixations (Martinez-Conde, Macknik, & Hubel, 2004): *tremor*, *microsaccades*, and *drifts*. Tremor is a small eye movement with high amplitude, and it is probably the result of imprecise muscle control (but the exact purpose is unclear). Drifts take the gaze slowly away from the fixated location and microsaccades bring it back. Microsaccades follow a straight line, while drifts have different curvature as visualized in Figure 1.4.

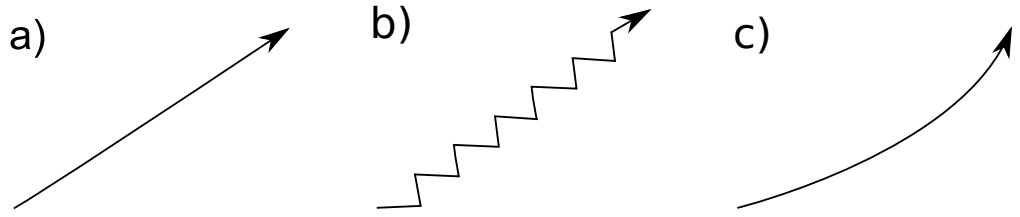


Figure 1.4: Schematic differences in curvature between microsaccades (a), tremor (b), and drifts (c).

A second type of eye movement is a *smooth pursuit*. Smooth pursuit eye movements are slower then saccades (velocity ranges from $10 - 30°/s$) and they occur when humans track some moving object. During smooth pursuit, eye position is continuously updated to look at the moving object. The conceptual difference between saccade and smooth pursuit is the fact that during most of the saccades, scene content is not processed at all (Dodge, 1900; Rayner, 1998) and visual

sensitivity even improves during smooth pursuit eye movement (Schutz, Braun, Kerzel, & Gegenfurtner, 2008).

Descriptive statistics of eye movements are summarized in Table 1.1.

Type	Duration (ms)	Amplitude	Velocity
Fixation	200–300	–	–
Saccade	30–80	4–20′	30–500′/s
Glissade	10–40	0.5–2′	20–140′/s
Smooth pursuit	–	–	10–30′/s
Microsaccade	10–30	10–40′	15–50°/s
Tremor	–	1′	20′/s
Drift	200–1000	1–60′	6–25′/s

Table 1.1: Descriptive statistics of different types of eye movements. Table adopted from Holmqvist et al. (2011)

1.4 Eye tracking

Eye movements can be tracked by specialized devices called *eye trackers*. In this section, we briefly describe the mechanism behind eye tracking. We describe a typical eye tracking measurement session and show some examples of eye tracking data.

1.4.1 Eye trackers

There is a long history of eye tracking mechanisms (Holmqvist et al., 2011). The first eye trackers were self-made by researchers, and they often required manual installation on the eye so they were rather uncomfortable for participants in experiments. In the mid-1970s, the first companies started to manufacture eye trackers. On the one hand, this lead to an increased interest in eye tracking; on the other hand, it lead to an unfortunate situation where researchers were forced to believe in the quality of both the hardware and software parts of eye trackers. There are three main types of eye tracking devices (Al-Rahayfeh & Faezipour, 2013):

- Tracking with attached sensor -– a small coil is embedded into a contact lense, and then the eye gaze is estimated based on the coil's orientation in a magnetic field (Robinson, 1963). This type of measurement is very precise; on the other hand, it is not very comfortable for the subjects of experiments.

- Sensor-based eye trackers – direction of the eye gaze is computed from electric signals from two pairs of electrodes placed around the eye. When the eyes move, the retina gets closer to one of the electrodes and the cornea to the other.

- Video-based technique – direction of the eye gaze is estimated using methods from computer vision. Usually, the eyes are recorded by a video camera and a small source of infrared light is directed at the pupil. This light source is partially reflected by the cornea. Therefore, the eye gaze direction is computed by measuring the difference between the reflected infrared light and the estimated center of the pupil.

Video-based eye trackers are currently the most commonly-used type of devices. Typically, the head should be constrained to view a scene from a predefined distance. Video-based eye trackers require a calibration procedure before the experiment, which usually consists of a small target randomly appearing on and moving across the screen while the participant looks directly at it.

Although eye trackers are usually placed in front of the monitor, there are also mobile eye trackers. The latter record both eye gaze and the scene viewed. This allows researchers to analyze eye movements in natural behavior: such as driving (Land & Lee, 1994) or making a sandwich (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003). See Hayhoe and Ballard (2005) for a review of eye tracking in natural tasks).

The typical sampling rate of eye trackers ranges from 10 Hz to 1000 Hz (Holmqvist et al., 2011). Therefore, for eye trackers with a low sampling rate, it is difficult to identify all events correctly, because the duration of the eye movements are below the sampling rate. For example, with an average fixation duration 200 ms and a sampling rate of 20 Hz, when the eye tracker samples eye position before the end

of the fixation, there is a window of 50 ms where we are not able to compute the end of the fixation correctly.

1.4.2 Process of eye tracking measurement – EyeLink II

Here we describe an example of measurement with one particular video-based eye tracker: EyeLink II (shown in Figure 1.5). The process is very similar for all

Figure 1.5: EyeLink II. Two cameras track the eyes and one camera in the center of the headband tracks four markers on the monitor. The picture was taken from `http://www.sr-research.com/EL_II.html`.

video-based eye trackers. Typical eye tracking measurement looks as follows. The participant is seated in front of the computer and the eye tracker is firmly positioned on his head. Both the front cameras are directed at the eye, so the pupil is in the middle of the projected image. Then, the threshold for pupil detection is adjusted so the pupil is correctly recognized as ellipse. There are four markers attached next to each corner of the monitor. In the center of the headband, there is an additional camera that tracks the position of the eye tracker with respect to the four markers. After adjusting the eye tracker, the participant's head is positioned on the chin rest at a fixed distance from the screen. This ensures the same perceived size of the presented stimuli for all participants.
Before the experiment, the eye tracker first needs to be calibrated. The calibra-

tion procedure works as follows. The target locations form a grid of nine points. A small circle appears on the central location, and the participant fixates on the center of the circle. When the gaze is stable (usually hundreds of milliseconds), the circle randomly changes its location to one of the other possible locations on the screen and the participant follows it with his gaze. For different levels of accuracy, more or fewer of the potential target locations can be used.

The calibration procedure is usually followed by validation to compute the measurement errors for the calibrated eye tracker. During validation, new target locations are computed and the participant's task is again to fixate on these locations as they randomly appear on the screen. Calibration error is expressed as average and a maximum distance between measured eye gaze and computed target location. Then, the eye with the lower error is selected for tracking (in case that binocular tracking is required, calibration error should be low for both eyes).

During the experiment, drift correction is used to verify whether or not calibration error increased during the experiment (for example, due to shifting of the headband). It can also serve as a fixation dot in the experiments.

Eyelink II can be controlled via low level libraries for a language C supplied by the manufacturer[1]. There are several wrappers in other languages to control Eyelink. The most notable one is library pylink for python. This library can be used with software such as PsychoPy (Peirce, 2007) or OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). There is also an Eyelink toolbox for MATLAB (Cornelissen, Peters, & Palmer, 2002) which is part of the Psychtoolbox extension for MATLAB (Brainard, 1997). Psychtoolbox is a collection of various functions that is often used in vision research. It allows for precise timing of the stimuli and measuring responses. It has an interface for various hardware (such as eye trackers). We used this toolbox in our research as well.

1.4.3 Eye tracking data

Data from Eyelink II is stored in internal binary format (extension .edf), which can be converted to text files (extension .asc) by the supplied utility `edf2asc`. An

[1] https://www.sr-support.com/

example of the text file is shown in the Listing 1.1. Each line with an eye gaze position begins with a time stamp; other lines include either custom messages (MSG), start/stop times for fixations (SFIX, EFIX) or saccades (SSACC,ESACC). The file contains information about calibrations, validations and drift corrections. It is then parsed by custom functions to obtain clean data. Detailed information about the file is listed in the manual for the eye tracker.

```
MSG      487908 trialStart
487912     373.7    387.2   1145.0 .
487916     373.6    387.1   1145.0 .
EFIX R   486996 487916  924        372.6    365.9     1131
SSACC R  487920
487920     375.7    387.1   1142.0 .
487924     377.8    387.1   1148.0 .
487928     393.6    396.4   1154.0 .
487932     411.7    400.5   1169.0 .
```

Listing 1.1: Example of Eylink II data file. Each line corresponds to one sample. Line starting with EFIX or SSACC denotes fixation and saccades (as computed by the implemented algorithm)

Although other manufacturers have different output file formats, the workflow and corresponding abstraction layer are similar thus allowing different software to cooperate with them.

1.4.4 Approaches to the analysis of eye tracking data

There are several approaches to the analysis of eye tracking data. One traditional approach is to identify fixation and saccades and then compare distribution of the descriptive statistics of the fixations (duration) or saccades (duration, orientation, peak velocity or latency). A typical example would be to compare the parameters of eye movements between two groups (for example, differences in the eye movements between patients with Alzheimer's disease and healthy controls; Molitor, Ko, & Ally, 2017) or within a group (comparing differences in eye movements when fatigue increases; Schleicher, Galley, Briest, & Galley, 2008).
Another possibility is to define areas of interest and compute dwell time for each AOI. Dwell time is computed as total time (or percentage) when a participant fix-

ated inside the particular AOI. This analysis is often used for static tasks, where the specification of AOI is easy. However, it is rarely used for dynamic tasks, because AOI have to be defined for each frame. The position of AOI could be computed for tasks where AOI move systematically. Still, for tasks such as movie watching, computations should be done per each frame. Recently, Papenmeier and Huff (2010) created a tool for matching eye movements to dynamic AOIs, which would help with the data preparation for the analysis.

Finally, for dynamic tasks, eye movements can be represented as scan patterns (sequences of spatio-temporal events) and compared using one the various methods described later in this paper (see Chapter 2).

1.4.5 Eye data quality

Despite all the progress in eye tracking, there is still a lot of noise in the system. Since there is no ground truth, we only get estimated locations of where people are looking. A lot of work has been done to compare the performance of eye trackers and estimate usual variances in eye tracking measurements per each eye tracker model. Holmqvist, Nyström, and Mulvey (2012) described the problem of noisiness for eye tracking data and the importance of having good quality data. For example when a scene is divided into AOIs and dwell time for each AOI is computed, adding 0.5° of spatial noise changes results to a very big extent[2]. Manufacturers often specify measurement error, but several studies have shown that after noise removal the average error was larger (Komogortsev & Khan, 2008; Zhang & Hornof, 2011). The measurement error of the eye trackers is described in terms of accuracy and precision. Low accuracy corresponds to systematic error; low precision corresponds to the more variable spatial distribution of the eye tracking measurements. The distinction between those two terms is visualized in Figure 1.6. It is important to know the precision and accuracy for the eye tracker used in the study. High accuracy is important for studies where similarity of eye movement is compared, while precision is important for studies where dwell time or fixation distribution are compared.

There are a lot of factors influencing eye tracking measurement errors, such as

[2]This spatial offset is considered as a very small error.

eye physiology (eyelashes, glasses, pupil diameter, etc.), and the skills of the eye tracking operator (Nyström, Andersson, Holmqvist, & van de Weijer, 2013); ethnicity, viewing distance, and head movement (Blignaut & Wium, 2014); as well as eye tracker design, recording environment, and experimental tasks (Holmqvist et al., 2012). Researchers should be aware of the possible sources of eye data variability and try to control for them.

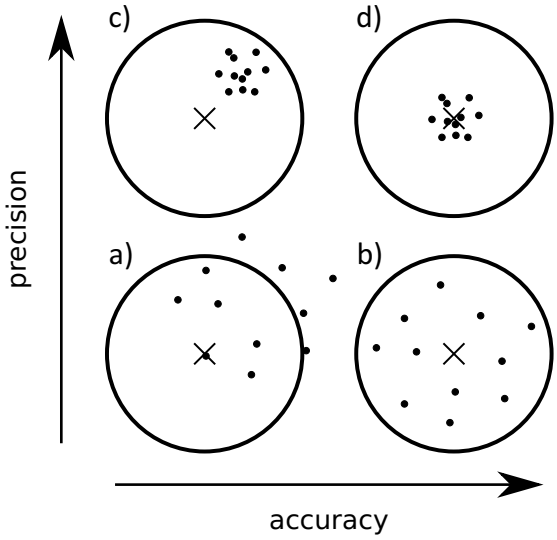


Figure 1.6: Distinction between precision and accuracy. Four schemas show data with low accuracy and precision (a), high accuracy and low precision (b), low accuracy and high precision (c), and high accuracy and precision (d).

1.5 Scan patterns

In every task where order of the fixations and saccades is important, eye movements are often represented as scan patterns. Scan pattern (or scan path) was originally defined by Noton and Stark (1971) as a visual pattern that the participants store in their memory when looking at some scene. When viewing a scene for the first time, the scan pattern is encoded into memory. During recognition, this pattern is compared with a new pattern when viewing the scene again. Although their theory was disputed (Foulsham & Kingstone, 2013), the term scan pattern is now used to represent eye movements as sequences of spatio-temporal data. In current literature, both the terms scan path and scan pattern are used.

However, we will use the term *scan pattern* to distinguish from the original meaning in Noton and Stark's theory (1971).

1.5.1 Scan pattern representation

There are two main approaches how to represent scan patterns. The first approach is called *event-based*. Using this type of representation, eye movements are classified into fixations and saccades, and scan pattern is therefore a temporal alignment of those events. Saccades and fixation are meaningful for visual perception, and thus this representation makes it easier to interpret scan patterns with respect to the given scene. In dynamic tasks, smooth pursuit is also common type of eye movement. Until recently, this type of eye movement was hard to detect in eye data due to its similarity to the set of fixations. If we used an event-based approach, interpretation of scan patterns would have a different meaning. An alternative approach would be to use raw data from an eye tracker instead. This approach can be used for finer relationships between scene content and the position of the eye gaze, which is particularly interesting in modelling.

1.5.2 Event extraction algorithms

Algorithms for converting raw data into events can be divided into the three groups based on the criteria they use for event detection: dispersion, velocity and acceleration (Duchowski, 2007). Dispersion-based algorithms detect fixations as sets of samples that are close to each other (e.g., 0.5°) for a minor time window (the time range is usually 80–150 ms). Everything else is treated as saccades. The most common algorithm is I-DT, which is implemented in many common eye trackers. Another dispersion algorithm for event detection models the event using Hidden Markov chains (Salvucci & Anderson, 1998). Velocity-based algorithms detect fixations based on a speed threshold, which is set after an initial check of velocities in the data (Salvucci & Goldberg, 2000). These algorithms are rarely used in the real data due to their sensitivity to noise. Acceleration-based algorithms use an approach similar to that of velocity-based algorithms, but they also detect saccade onset and offset to better start and end the fixation.

For eye trackers with low sampling rates, there are several approaches to detecting

smooth pursuit using Bayesian clustering (Tafaj, Kasneci, Rosenstiel, & Bogdan, 2012) or Bayesian decision theory (Santini, Fuhl, Kübler, & Kasneci, 2016). For high sampling eye trackers, a new algorithm for detecting smooth pursuit was developed (Larsson, Nyström, Andersson, & Stridh, 2015) and recent comparisons show its high efficiency (Andersson et al., n.d.). Another recent approach to smooth pursuit detection is based on extracting smooth pursuit from samples using multiple observers (Agtzidis, Startsev, & Dorr, 2016). Smooth pursuit detection in eye tracking data from mobile eye trackers is still an unsolved question.

1.5.3 Scan patterns in dynamic and static tasks

Although every eye movement can be represented as scan patterns, this representation is not ideal. In static scenes, the temporal ordering of fixated locations is highly variable. Therefore, temporal alignment of two scan patterns would lead to high dissimilarity. In the dynamic tasks, participants' gaze is more similar. When watching movies, participants start looking at similar places and similarity decreases over time as they drift apart. However, after each cut in the movie, eye movements synchronize (Smith & Mital, 2013).

To get an idea about the variability of scan patterns, we present scan patterns from static and dynamic tasks (Figure 1.7). In the left part of figure, there are three scan patterns from one subject searching for a Gabor patch target in Gaussian noise in a repeated presentation of the same trial. In the right part of figure, we show scan patterns from three subjects viewing the same movie. There are visual differences between the scan patterns for each task. For the visual search, there are plateaus that represent fixations alternated by saccades. There is no general pattern. This particular task used static noise, which was hard to distinguish from other randomly-generated noises. Therefore, subjects did not recognize the repetition. The scan patterns from the movie contained smooth pursuit eye movements as well, so the fixations are not clearly detectable as in the visual search task. We could not present identical movies repeatedly, because participants would recognize the repetition and then possibly focus on different semantic content in the movie. It is important to note that we used different eye trackers with different accuracy and precision (Eyelink 1000 for the visual search

task and Eye Follower for the movie presentation), so direct comparison of both plots would be misleading.

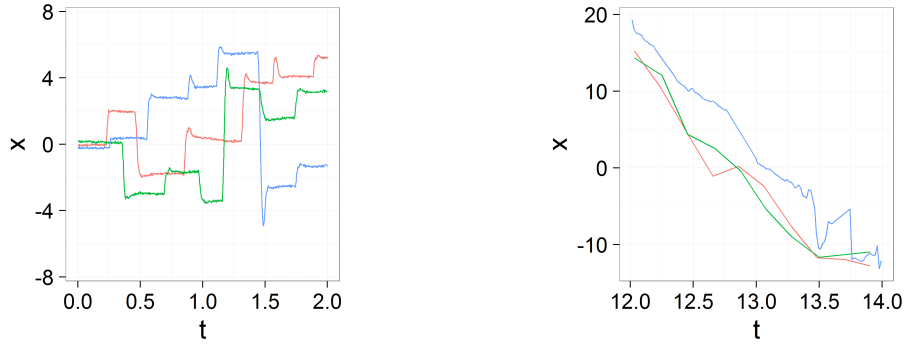


Figure 1.7: X coordinate in time in a static task (left) and in a dynamic task (right)

1.6 Multiple Object Tracking

Eye tracking is used for various tasks. In this dissertation, we used one task that has several parameters suitable for studying eye movements. This task is called *Multiple Object Tracking*. Multiple Object Tracking (MOT; Pylyshyn & Storm, 1988) was developed for studying divided attention. In this task, the participant's goal is to track several moving objects (targets) among other objects (distractors). A typical trial looks as follows (depicted in Figure 1.8). First, all objects are displayed for couple of seconds and targets are highlighted (by changing color or flashing). Then targets change color (or stop flashing) to become indistinguishable from the distractors and they start to move around the display for a couple of seconds. Objects bounce off each other and after the movement stops, the participant's task is to select tracked objects (or he/she is sometimes queried about a specific object).

1.6.1 Parameters influencing tracking accuracy

In general, there is a lot of individual variability in terms of tracking capacity (Oksama & Hyönä, 2004). Tracking is demanding on resources, and thus participants mainly use positional information. When the objects change shape or

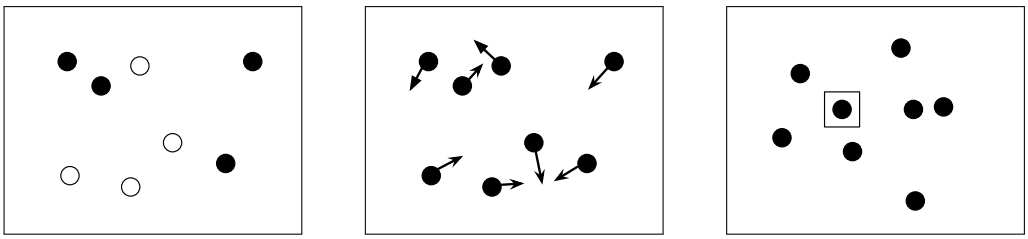


Figure 1.8: Example of the trial in MOT. Description is in the text.

color during the tracking, this information is usually not retained (Bahrami, 2003; Saiki, 2003). Also, the identity of the targets is not retained during the tracking (Pylyshyn, 2004; Treisman & Zhang, 2006).

Typically, tracking accuracy is dependent on the number of parameters the task has. Participants are usually able to track four out of eight objects (Intriligator & Cavanagh, 2001; Yantis, 1992). With an increasing number of distractors, tracking becomes harder (Alvarez & Franconeri, 2007) and similarly it is harder to track more targets at the same time (Pylyshyn & Storm, 1988). The capacity to track four objects was linked to the capacity of visual working memory (Luck & Vogel, 1997)). However, there is an open discussion as to whether this limitation is due to the number of available slots in working memory or due to the limits of the shared pool of resources (Brady, Konkle, Alvarez, & Oliva, 2008; Suchow, Fougnie, Brady, & Alvarez, 2014; Pylyshyn, 1989; Cowan, 2001). In MOT tasks in particular, the tracking capacity can go up to seven or eight targets at once (Alvarez & Franconeri, 2007).

More predictable trajectories for objects increase tracking accuracy (Howe & Holcombe, 2012). When targets disappeared while moving, tracking became harder when they re-appeared several milliseconds later at the extrapolated position rather than at the original position (Keane & Pylyshyn, 2006). This showed that participants did not use motion information for tracking but positional information only. However, Fencsik, Klieger, and Horowitz (2007) showed that the motion is extrapolated when only two targets are tracked. Therefore, the lack of motion extrapolation could be the result of limited attention resources. The extent and potential benefit of motion extrapolation is still an open question; i.e. as to whether it helps tracking (St.Clair, 2010; Howe & Holcombe, 2012) or not

(Vul, Frank, Tenenbaum, & Alvarez, 2009).

An important parameter for tracking accuracy is the velocity of the objects. Higher velocities lead to decreased tracking performance (Pylyshyn & Storm, 1988; Liu et al., 2005). At very low velocities, participants could track eight objects at a time, while for some velocities they could only track one object (Alvarez & Franconeri, 2007).

Similarly, when the objects are packed more closely to each other, tracking gets harder (Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008; Shim, Alvarez, & Jiang, 2008). The shape of objects also affects tracking. Usually, simple geometrical objects are used as stimuli (circles, squares), but when there is not a clear distinction between the targets and the distractors; such as tracking one end of a straight line while the other end is used as distractor. In such cases, tracking accuracy decreases. When the targets and distractors are connected by a line forming a dumbbell, tracking again gets easier (Scholl, Pylyshyn, & Feldman, 2001).

It seems that visual systems have individual resources for tracking in each hemifield (Alvarez, 2005). Therefore, participants are not able to track four targets but rather two targets in each hemifield separately. This leads to the alteration of this paradigm, where two objects circle around each other similar to planetary movement (one target and one distractor) and each item in the pair is located in the one corner of the display (Franconeri, Jonathan, & Scimeca, 2010). This modification of MOT allows for more precise control of objects' velocity (Tombu & Seiffert, 2011) or distance between a target and a distractor (Meyerhoff, Papenmeier, Jahn, & Huff, 2015).

Finally, the size of the virtual arena in which objects move limits tracking accuracy as well. This is probably related to the spatial resolution of attention (Intriligator & Cavanagh, 2001).

1.6.2 Mechanism behind tracking multiple objects

There are several theories explaining the ability to track multiple objects at the same time. As described in Scimeca and Franconeri (2015), there are three main theories that explain tracking performance. First, there could be one attentional spotlight (attention allocated to a particular area in the scene) that shifts between

targets (Pylyshyn & Storm, 1988). Therefore, the ability to track the targets is limited by the temporal resolution of attention, because the attentional spotlight cannot switch between targets fast enough. We will call this model the *switching model*. Second, there can be multiple attentional spotlights where each of them moves with the allocated target (Cavanagh & Alvarez, 2005), this model will be called the *multifocal attention model*. Third, as proposed in the original paper (Pylyshyn & Storm, 1988), targets can be tracked preattentively by assigning pointers (called FINST – FIngres of INSTantiation) and only update the pointer location during tracking. This updating is done without attention, and this is the main distinction from the multifocal attention model. We will call this the *FINST model*. Both multifocal attention and FINST model are limited by spatial resolution. Both the multifocal attention and FINST models are limited by spatial resolution. The final model explaining tracking of multiple targets is from Yantis (1992). He proposed that a visual system creates one virtual object from the targets (having targets as vertices), and this object is tracked by a single attention spotlight. We will call this the *grouping model*. Performance in this case is limited by the visual system's shape recognition ability.

It would be promising to select a valid model by exploiting the limitations that constrain each of them. However, it is hard to distinguish between the different models, because all differences in tracking accuracy can be explained by three types of competitive resources for the visual system: spatial, temporal and shape of the objects in the scene (Scimeca & Franconeri, 2015). *Spatial* resources are related to cortical maps that represent regions in the visual field. When two regions overlap, the visual system's detection capabilities are reduced. This phenomenon is called crowding and it is described in Section 1.6.3. *Temporal* resources are related to the speed of the processing mechanism that shifts among the tracked targets over time. *Shape* limits are related to the shape recognition system. When changing a task's parameters, changes in performance could be explained by each of these three types of resource limitations. For example, when objects move faster, due to spatial limitations, there are more situations when crowding occurs. Due to temporal limitations, higher speed would lead to an outdated position when the processing resource shifts among the targets. Finally, due to

shape limitations, higher speed would lead to difficulties in updating the shape. One way to distinguish between those models and competitive resources would be to study eye movement. Eye movement models are described in the next section.

1.6.3 Eye movements in MOT

In the first MOT experiments (Pylyshyn & Storm, 1988), participants were told to look at the center of the screen to track the object with attention only. However, recent studies have shown that it is beneficial to shift one's gaze during the task. Eye movements in the MOT are influenced by both top-down and bottom-up sources. Participants plan their eye movements to track targets successfully, but their plans are influenced by the local configuration of the objects on the screen. Eye tracking strategies correspond to the theories behind tracking objects. This relationship is not surprising, because although eye movements and attention can be dissociated, they are synchronized during normal cognitive load (Posner et al., 1980).

There are several approaches to explaining eye gaze position during MOT trials. The current approaches to modeling eye movement in MOT are mostly relatively simple. The first group is strategies that predict eye gaze position based on the configuration of the objects in the frame. We will refer to those strategies as *analytical strategies.* There is also one model predicting eye movements using Bayesian inference and short-term memory. Analytical strategies can be related directly to the models of MOT.

Analytical strategies

Analytical strategies in MOT were first described by (Fehd & Seiffert, 2008). They measured eye gaze during tracking and manually classified strategies used by participants into three groups. The first group did not move their eyes and tracked targets with attention only. The second group followed the general motion of the targets, which is in line with the grouping model for MOT (*center-looking*). The last group of participants switched their eye gaze rapidly between targets. In their follow-up study, they varied the speed of objects and showed that a preference for center-looking over switching between targets is not a result of avoiding

saccades due to saccadic suppression (Bridgeman, Hendry, & Stark, 1975; Burr, Morrone, & Ross, 1994), but comes rather from a general preference for this type of strategy. Fehd (2009) also proposed a centroid-target-centroid strategy in which participants switch between targets. This leads to better tracking accuracy. Participants following this strategy switch back and forth between centroid and individual targets.

Preference for a centroid strategy was supported in work from Zelinsky and Neider (2008).In their study, they used computer models of sharks swimming in a 3D aquarium. They also identified different tracking strategies similar to Fehd and Seiffert (2008). When subjects tracked two or more targets, they fixated near the centroid. However, in the case of four targets, participants spent more time fixating on individual targets than fixating on the centroid. In their follow-up study (Zelinsky & Todor, 2010), they introduced the term rescue saccades, which corrects for situations in which some target might get lost due to occlusion.

The above-mentioned studies employed target positions only. Landry, Sheridan, and Yufik (2001) studied plane tracking, and they found out that participants spent more time on the planes that were about to collide with each other than on planes with no chance of collision. Lukavský (2013). reported a similar finding. He proposed a model predicting eye gaze to the averaged position of the objects, but biased to the targets closer to the distractors, to prevent the chance of swapping the targets with distractors.

This phenomenon is known as *crowding* and it limits human perception on the periphery (Levi, 2008). Crowding is defined as deleterious influence of nearby contours on visual discrimination (Levi, Song, & Pelli, 2007). It is illustrated in Figure 1.9. While fixating on a cross, it is easy to identify the letter A on the left but harder to identify the equally distant letter A on the right, which is crowded by the other letters. Lukavský (2013) also made a distinction between

Figure 1.9: Example of the crowding phenomenon. When fixating on the central cross, letter A on the left can be easily identified, while the equally distant letter A on the right cannot be identified because of the surrounding letters.

the center of the convex hull of the targets and the centroid of the virtual object. This went unnoticed in the previous models. This strategy that minimized the chance of crowding was extended in Děchtěrenko and Lukavský (2014) showing that it is consistently better even for trials with larger numbers of distractors (and therefore a higher chance of crowding).

Recently, Lukavský and Děchtěrenko (2016) showed that participants' eye movements lag approximately 100 ms behind the content of the scene. In their experiment, they repeatedly presented trials in forward and backward conditions (object followed identical trajectories as in forward condition, but with reversed time coordinate). They reversed the scan patterns to the same time coordinates as in the forward condition and shifted the scan patterns across the time dimension to obtain maximum similarity. They found consistent effect across four experiments independent from the predictability of the object movement or workload. Incorporating this factor into the models better explained the human eye gaze.

Bayesian model of eye movements in MOT

An interesting model for predicting eye movements using double-layered Bayesian architecture was proposed by Colas, Flacher, Tanner, Bessière, and Girard (2009). The first layer represented the probability of object being at the particular position (occupancy grid), and the second layer represented the remembered positions of the targets across the occupancy grid (memory layer). Using Bayesian inference, both layers were updated for each time step, and they worked with three models using parts of this two-layered architecture. First, the Constant model served as a baseline, and it predicted the eye gaze position to be constant throughout the trial (therefore, it did not use the memory layer). Second, the Target positions model inferred the eye gaze position using both layers; usually predicting the eye gaze in a weighted sum of the locations of the targets. The third model was the Uncertainty model which extended the previous model by adding uncertainty about the targets. The Uncertainty model outperformed the other two.

The relationship between models accumulating information from the beginning

of the trial to the analytical strategies working with the current frame only is unknown. Lux (2014) implemented models from Colas et al. (2009) and compared their performance to models used by Děchtěrenko and Lukavský (2014). Predicted scan patterns from analytical strategies outperformed scan patterns from Bayesian models (Normalized Scanpath Saliency was used for comparison). These results showed that eye movements may be planned using simple rules. There were two differences between the original Bayesian models and the one from Lux (2014). They could be the source of the difference in performance. First, the original models used log-complex retinotopic maps, while the replicated model used linear versions instead. Second, the original model used a different formula for quantifying uncertainty. Therefore, further testing should be performed to state the differences in prediction strength.

1.6.4 MOT as a playground for studying eye movements

So far, researchers have been interested in how one could explain eye movements in a MOT task. However, we could exploit MOT tasks in a different way. Participants usually do not recognize repetition. Ogawa, Watanabe, and Yagi (2009) showed that even with 15 identical repetitions of an identical trial, the recognition rate was only 22%–31%. Lukavský (2013) showed that for four repeated presentations of an identical trial, only 24% of participants claimed they recognized the repetition. When they were tested directly, they were able to recognize repetition in 47% of the trials. On the other hand, they also showed high false alarm rates (44%). Therefore, their discrimination capabilities were low.

Previous results also show that scan patterns from the repeated presentations of the identical trials show high similarity. The most similar scan patterns are two patterns from the same subject and the same trial; followed by scan patterns from different subjects and the same trial, the same subject and a different trial, and the least similar were scan patterns from different subjects and different trial (Lukavský, 2013).

Difficulty in recognizing repetition, taken together with the high similarity of intra subject measurement, creates a powerful playground for studying the properties of scan patterns. The nature of the task allows researchers to modify trajectories

systematically and study how this behavior influences the shape of scan patterns. This is a great advantage of this task compared to using complex stimuli such as movie presentation. We could not easily alter the content of the movie clip without participants noticing, but we can change the trajectories in MOT in various ways. For example, we could rotate all object trajectories around the center or flip them around one of the axes. This task could therefore serve as a great tool for methodological experiments.

1.7 Purpose of this thesis

This thesis focuses on the statistical comparison of scan patterns in dynamic tasks. For this purpose, we employed the Multiple Object Tracking paradigm. This allows us to study intra-subject variability of scan patterns and thus reveal the amount of inherent noise present in the data. The main advantage of using MOT for this purpose is the precise formalization of the movement of the objects. Each of them follows a trajectory, which can be altered without any loss of meaning (e.g. adding Gaussian spatial noise to the trajectory would result in more Brownian-like motion, which does not change the interpretation of the task). In particular, we focused on the metrics that are generally used for comparing scan patterns and evaluated their performance on the scan patterns with added noise.

In the second part of the thesis, we focus on the problem of statistical testing of differences between groups of scan patterns. This is a general question applicable to any two groups of time series.

In the third part, we show how scan patterns in MOT could be predicted by the feed-forward neural networks.

2. Comparison of scan patterns

In every research study involving scan patterns, there is a need to quantify the similarity of scan patterns. There are several metrics which serve this purpose. In this section, we will give an overview of those metrics and divide them into two groups based on whether they used scan patterns as successions of events in time or whether they used raw samples instead. One general problem with metrics for scan pattern comparison is that it is unclear how we can translate the estimates of similarity from one metric to another and how these estimates correspond to differences in scan patterns. In this chapter, we first describe metrics that are commonly used for comparing scan patterns. Then, in four simulations, we will explore the relationships between several chosen metrics and provide guidelines for further interpretation.

2.1 Similarity versus coherence

In this and subsequent chapters, we will be comparing scan patterns. We will use several interchangeable terms. For the two scan patterns, we will use the terms *similarity* and *distance* where the former denotes the similarity of the two scan patterns and the latter the distance between two scan patterns on the scale of the given metric. For groups of scan patterns, we will use the term *coherence* for the overall similarity of groups of scan patterns and *average distance* for the overall distance. Similarity and distance are inverse terms and therefore some of the metrics have different meanings in their interpretations.

2.2 Event-based methods

Researchers usually represent eye data as scan patterns in static tasks where the events could be extracted from the raw data. Therefore, the majority of metrics work with event-based scan patterns. Here, we describe several of the common event-based metrics.

2.2.1 Levenshtein metric

One of the most typical metrics is the *Levenshtein distance metric* (Levenshtein, 1966), or string edit distance. This metric employs Levenshtein distance used for computing similarity between two strings. The idea behind this metric is simple. The presented scene is divided into areas of interest (AOI) that can be either grid-like or semantic-like. Grid-like AOIs divide a scene into small rectangular areas and scene content is not taken into account when defining AOIs. The number of AOIs can influence the similarity of two scan patterns (Kocián, 2014). Semantic-like AOIs divide the scene based on its content. For example, for analysis of scan patterns on webpages, one AOI can be defined as a search bar, another as a menu, etc. When using this approach, it is more sensible to interpret the scan patterns with respect to individual AOIs. Examples of grid-like and semantic-like AOI are visualized in Figure 2.1.

For each AOI, a character is assigned to that area and a scan pattern is described as a string. After that, Levenshtein distance is used to compute the distance between strings as number of edits, removals or inserts to modify one string to another. The Levenshtein metric can be used for scan patterns with raw samples as well. We simply encode each sample with a letter and compute distance between those long strings. This representation captures a very rough estimate of spatio-temporal similarity. However, for large AOIs, lots of variance in fixations would be binned into the same region.

The main advantage of this metric is its simplicity, because algorithms for string distance are simple and they are often implemented in various libraries for different programming languages. Also, this metric was used in the first studies regarding scan patterns (Brandt & Stark, 1997), and therefore researchers could compare the results to the previous studies. The main disadvantages of Levenshtein distance is the problem with grid borders. When two fixations are spatially close to each other, but are in different AOIs, they are treated as dissimilar.

2.2.2 ScanMatch metric

Another metric for comparing scan patterns is ScanMatch (Cristino, Mathôt, Theeuwes, & Gilchrist, 2010). This metric extends the idea of finding the best

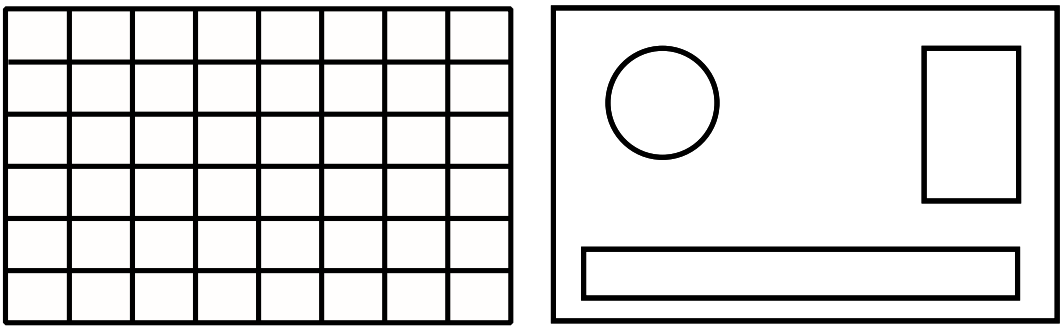


Figure 2.1: Areas of interest: grid-like (left) or semantic-like (right).

global alignment of scan patterns similar to the Levenshtein metric. However, in contrast to the Levenhstein metric, ScanMatch searches for the optimal match using the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970) used for aligning sequences of genomes. This algorithm uses a substitution matrix for inconsistencies between scan patterns. This matrix is usually defined as simple Euclidean distance between AOIs (on the contrary, the Levenshtein metric penalized inconsistencies using a constant value for all mismatches), but it could be also defined by color or semantic similarity. It also uses a gap penalty, which adds a penalization for missing parts of scan patterns. An algorithm with a small gap penalty would try to find local alignments of scan patterns, while a large penalty would lead to the alignment of entire scan patterns. ScanMatch also encodes the temporal duration of fixations. Samples are binned by temporal dimension and each bin is encoded with the corresponding number of letters. For example, when the bin size is 50 ms and fixation duration in AOI with letter x is 120 ms, we encode the bin size as *xxx*.

Cristino et al. (2010) showed that ScanMatch outperformed Levenshtein distance for artificial data. For static tasks, ScanMatch can be used for raw samples similar to Levenshtein distance.

2.2.3 Mannan's metric

Another metric used for scan patterns is Mannan's metric (Mannan, Ruddock, & Wooding, 1995, 1996, 1997). Average distance is computed by calculating the distance between all fixations in one scan pattern and the closest distance in the second scan pattern. This distance is then compared to the randomized scan

pattern, and final similarity is expressed as

$$I_s = (1 - \frac{D}{D_r})$$

where D is computed as

$$D^2 = \frac{n_1 \sum_{j=1}^{n_2} d_{2j}^2}{2n_1 n_2 (a^2 + b^2)} + \frac{n_2 \sum_{i=1}^{n_1} d_{1i}^2}{2n_1 n_2 (a^2 + b^2)}$$

where n_1 and n_2 are numbers of fixation in the scan patterns, d_{1i} is the distance between ith fixation in the first scan pattern and the closest fixation in the second (and vice versa for d_{2j}, a and b are the lengths of the arena (or presented stimuli) and D_r is a randomized scan pattern. Index I_s ranges from 0 (random scan path) to 100 (identity).

Mannan's metric does not také the temporal order of the fixations into account. This could compensate for the high noisiness in the static tasks, but it would also claim that the two scan patterns are identical, when one is actually the reverse copy of the second. Also, scan patterns with different numbers of fixations (i.e. one scan pattern has more fixations than the other one) could easily be treated as similar (Le Meur & Baccino, 2013).

2.2.4 MultiMatch

Jarodzka, Holmqvist, and Nyström (2010) proposed a comparison method, which compares the similarity of scan patterns across several dimensions at once. First, saccades in the scan patterns are represented as vectors (the curvature is simplified to a straight line) and fixations represent the vectors start/stop points (fixation duration is also part of the representation). Therefore, a scan pattern is represented as an ordered set of $n-1$ vectors, where n is the number of fixations in the scan pattern. The scan pattern is then simplified by applying two steps repeatedly. First, consecutive small saccades $u_1, \ldots u_k$ with amplitude smaller than a threshold T_{amp} are replaced by averaged vector u'. Second, consecutive saccades with direction lower than threshold T_θ are also replaced by one vector. The fixation durations are merged appropriately. This simplification keeps the similarity on a global level.

For each vector in each scan pattern (consisting of n and m vectors respectively),

similarity of each element is computed with all other elements – $M(i, j)$ for i-th vector from the first scan pattern and j-th vector in the second scan pattern. The vectors' adjacency is captured in the matrix $A(k, l)$, where $k, l = 1, 2, \ldots, mn$ (each combination of vectors) and each connection is associated with the appropriate weight $M(i, j)$.The adjacency matrix forms a graph in which the shortest path represents the temporal alignment of the two scan patterns. For this representation, we can compare several parameters of the scan patterns at once: such as difference in shape between vectors $(u_i - u_j)$, difference in amplitude between vectors $(\|u_i - u_j\|)$, distance between fixations, differences in direction between vectors and difference in duration between fixations (Jarodzka et al., 2010).
This method does not need to specify AOIs as is done in string edit distances, and it captures different dimensions of similarity all at once. On the other hand, this method could not be easily extended for dynamic tasks with smooth pursuit or capture the coherence of whole groups of scan patterns.

2.2.5 Recurrence quantification analysis

Another approach to the scan pattern comparison is to view eye movements as dynamic systems. Anderson, Bischof, Laidlaw, Risko, and Kingstone (2013) analyzed scan patterns using recurrence quantification analysis, which was successfully used for describing complex dynamic systems (e.g. Marwan & Kurths, 2002). For a sequence of fixations f_i, the fixations are recurrent if they are close to each other. The proximity can be specified as

$$r_{ij} = \begin{cases} 1, & d(f_i, f_j) \leq \rho \\ 0, & \text{otherwise} \end{cases}$$

where d is the distance metric (such as Euclidean distance) and ρ is a parameter specifying the threshold for proximity. Therefore, we can visualize similarity using a recurrence plot, which shows dots on the coordinates corresponding to the recurrent fixation. The recurrence measure could be defined as $REC = \frac{2R}{n(n-1)}$ where n is number of fixations and R is defined as $R = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} r_{ij}$ that corresponds to the sum of the upper triangle of the matrix. They also defined several other measures based on the recurrence analysis; such as determinism

(representation of a repeating scan pattern in the diagram), laminarity (rescanning of an area that had been previously scanned, albeit briefly) or center of recurrence mass (the position where the recurrent points are situated in time). This method is used more often for eye movements that are not represented as scan patterns (Anderson, Laidlaw, Bischof, & Kingstone, 2012).

2.2.6 Earth mover's distance

Earth mover's distance[1] is a measure typically used to compare the distance between two probability distributions. Dempere-Marco, Hu, Ellis, Hansell, and Yang (2006) used this measure to compare similarity between two scan patterns. Scan patterns are treated as two sets of fixations where one set of fixations can be visualized as holes and the other as a dirt needed to fill the holes. The fixation duration corresponds to the size of the holes. Formally, we need to minimize the transportation function

$$cost(F_i, F_2, R) = \sum_{i=1}^{m} \sum_{j=1}^{n} d(f_{1i}, f_{2j}) r_{ij}$$

where F_1 and F_2 are sets of fixations (with f_{1i} and f_{2j} as individual fixations), n and m denote number of the fixations, d is a distance measure and R is the overall flow between the distributions (with r_{ij} and denotes a flow between the distributions. EMD distance is then defined as

$$EMD(F_1, F_2) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d(f_{1i}, f_{2j}) r_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} r_{ij}}$$

2.3 Raw sample-based methods

An alternative approach to representation of scan patterns is using raw samples instead of events. In the case of dynamic tasks, this is a fruitful alternative because of the prevalence of smooth pursuit movements. We are usually interested in the coherence of groups of scan patterns. Therefore, some measure of spatio-temporal distance between scan patterns can answer this question. In literature, the raw-sample based method is often denoted as saliency-based measures, because researchers often compute a saliency map based on the fixations (Le Meur

[1] Also known as Wasserstein metric

& Baccino, 2013).

The comparison methods are usually divided into ones that compare two maps both to each other and to methods that compare sets of fixated locations to a salience map[2] (Le Meur & Baccino, 2013). Because scan pattern can be converted to the saliency map by convolving with Gaussian filter, we do not distinguish between the classifications.

2.3.1 Saliency map versus spatio-temporal map

A saliency map is another representation of the scan pattern. The saliency map represents areas in the scene where the participant fixated during the task. The term is taken from saliency defined by Itti, Koch, and Niebur (1998). Those fixation points are usually smoothed by an isotropic bi-dimensional Gaussian function (Le Meur, Le Callet, Barba, & Thoreau, 2006). Due to smoothing, two scan patterns slightly shifted in one of the spatial coordinates are treated as similar. This approach does not take temporal order of fixation into account, but this could be fixed by extending the saliency maps into a 3D variant, denoted as spatio-temporal fixation maps. Convolving scan patterns with a spatio-temporal Gaussian filter preserves similarity for time scale as well. Therefore, identical scan patterns shifted in space or in time are treated as very similar up to a certain degree defined by the properties of the Gaussian filter.

Both saliency maps and spatio-temporal fixation maps can consist of several scan patterns. In that case, each scan pattern is convolved with Gaussian filters separately. Then the spatio-temporal maps are summed into one map and the map is normalized into a 0–1 range. In such a case, the parts of the spatio-temporal fixation map, in which several scan patterns are similar, would have higher values. The process of creating saliency maps is visualized in Figure 2.2.

2.3.2 Correlation-based measures

Correlation-based measures (Jost, Ouerhani, Wartburg, Müri, & Hügli, 2005; Le Meur et al., 2006; Rajashekar, van der Linde, Bovik, & Cormack, 2008) can be

[2]fixated locations mean raw positions in this case

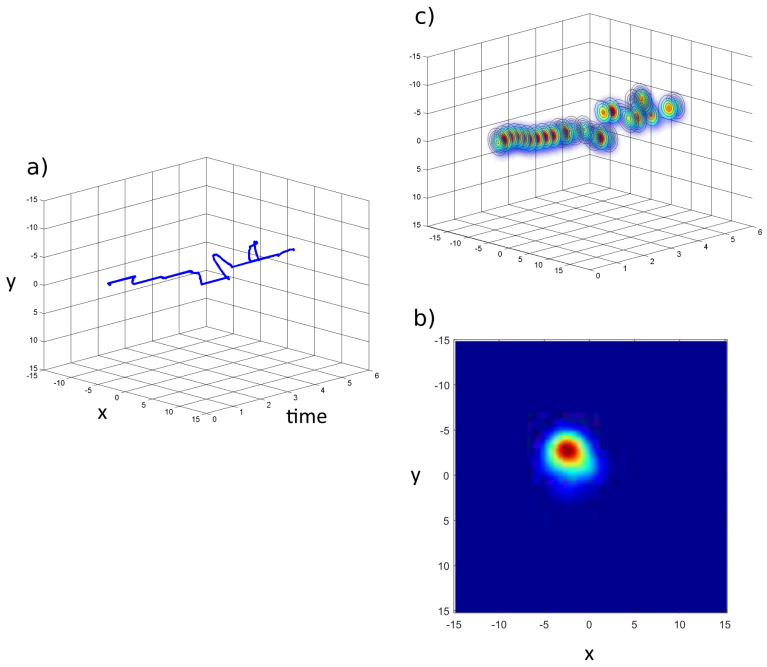


Figure 2.2: Saliency map versus spatio-temporal map. Scan pattern (a) is convolved with a spatio-temporal Gaussian filter creating a spatio-temporal fixation map (b) or with a bi-dimensional Gaussian filter creating a saliency map (c).

used to evaluate the similarity of two maps or to compare the similarity of a set of fixation points with a saliency map. Originally, it was used for saliency maps, but it has recently been used for spatio-temporal maps as well (Lukavský & Děchtěrenko, 2016; Děchtěrenko, Lukavský, & Holmqvist, 2017). The similarity of fixation maps is computed using a Pearson correlation coefficient. Therefore, similarity ranges from -1 (two maps are completely opposite) to 1 (identical saliency maps). Spearman's rank correlation coefficient offers another possibility for measuring the correlation between saliency maps (Toet, 2011).This metric computes the similarity of saliency maps instead of the distance. The main advantage of this measure is its intuitive interpretation. In this thesis, we show the extension of correlation-based measures to correlation distance that we used in other projects (Děchtěrenko et al., 2017; Lukavský & Děchtěrenko, 2016). Correlation distance is computed as $CD = 1 - r(M_1, M_2)$, where r is a Pearson correlation coefficient and M_1 and M_2 are two maps (either two or three dimensional) containing either one or more scan patterns. For maps with normalized values that range from -1 to 1, the Pearson correlation coefficient also ranges from

-1 (the two maps are completely opposite) to 1 (identical saliency maps). However, for the maps created using a convolving scan pattern with a Gaussian filter, the correlation coefficient reaches zero only occasionally. Due to the nature of fixation maps, the correlation coefficient can occasionally be less than 0. Therefore, for negative correlation coefficients, the CD metric is set to one. Consequently, the values of the CD metric can range from 0 (absolute correspondence) to 1 (completely different trajectories).

The main advantages of the CD metric are the limited range [0; +1] and the intuitive evaluation of results in comparison to other metrics mentioned later.

2.3.3 Normalized Scanpath Saliency

Another metric that uses saliency maps is Normalized Scanpath Saliency (NSS Peters, Iyer, Itti, & Koch, 2005). This metric is typically used for comparison of a saliency map and a set of fixations (Le Meur & Baccino, 2013), but it can also be used for computation of the similarity of scan patterns. Again, the use of this metric was extended to spatio-temporal maps as well (Dorr et al., 2010; Lukavský, 2013).

This metric takes saliency map F and normalizes it using z-transformation to

$$Z_F = \frac{F - mean(F)}{std(F)}$$

The NSS value is then computed as

$$NSS = \frac{1}{n} \sum_{i=1}^{n} Z_F(v_i)$$

where n is the number of samples in the scan pattern, v_i is i-th sample of the scan pattern and $Z_F(v_i)$ is the corresponding value from the map. For two-dimensional case, the computation can be visualized as an average altitude of the path (scan pattern) alongside some hills (saliency map).

2.3.4 Percentile metric

Peters and Itti (2008) created a metric that expresses similarity as an average ratio between a number of locations lower than the position on the saliency map

and with values less than the saliency value at the point specified by the scan path. Formally, it is defined as

$$P = \frac{1}{n} \sum_{i=1}^{n} 100 \times \frac{|\{x \in X : F(x) < F(v_i)\}|}{|F|}$$

where n is the number of samples in the scan pattern, X is set of all locations in saliency map, x is a two-dimensional vector of locations on the map, v_i is i-th location of the scan pattern and $|\cdot|$ denotes set size. Again, this method could be naturally extended to spatio-temporal maps as well. The percentile metric ranges from 50% (chance level) to 100% (absolute correspondence).

2.3.5 The Kullback–Leibler divergence

In general, the Kullback–Leibler divergence (KL-divergence) is commonly used in information theory to measure overall dissimilarity between two probability density functions. For two discrete distributions R and P, the KL-divergence is defined as

$$KL(R, P) = \sum_{k} p_k log \frac{r_k}{p_k}$$

where p_k and r_k are probability discrete functions (both p_k and r_k sums to 1) and if $p_k > 0$ for any k than $r_k > 0$ as well.

This method was used by Rajashekar, Cormack, and Bovik (2004) and Tatler, Baddeley, and Gilchrist (2005) for the comparison of saliency maps. Both saliency maps are transformed to probability density functions by adding small, non-zero values to all locations on the map to avoid division by zero and by dividing each position by the sum of the map.

For two identical maps, the KL-divergence equals zero and it increases as the saliency maps differs. It does not have an upper limit and it is also not a distance because it does not satisfy the triangle inequality.

2.3.6 Receiver operating characteristics

A final measure that works with the saliency map is based on the Receiver Operator Characteristic (ROC) analysis from signal detection theory (Green & Swets, 1966). This method takes one saliency map as the basic truth and the other one

as a prediction. Then it thresholds both saliency maps for different levels. From the pairs of values for basic truth and prediction, it computes values for true positive (TP), true negative (TN), false positive (FP) and false negative (FN). From those values, sensitivity (TP/(TP + FN)) and specificity (FP/(TP + FN)) are computed. For each threshold level, they are plotted on the graph. The curve connecting different sensitivity and specificity values is called the Receiver Operating Characteristics, and accuracy of classification is measured using the area below the curve, where the chance level equals 0.5 and a perfect match equals 1.

2.3.7 Fréchet distance

Fréchet distance (Fréchet, 1906) is a metric measuring the distance between two curves. It takes order of points into account. Therefore, it is used more often than well-known Hausdorf distance. It can be intuitively imagined as follows. A man is walking a dog, and both of them follow their own trajectory. They can vary their speeds, but they cannot go back on the curve. Fréchet distance is the minimum length of leash needed for the man to walk the dog (both man and dog are walking at their own speed and can wait for each other). Formally, if $f : [a, b] \rightarrow R^2$ and $g : [a, b] \rightarrow R^2$ are curves in space[3], then Fréchet distance is defined as $\delta_F(f, g) = \inf \alpha, \beta \max_{t \in [0,1]} d(f(\alpha(t), g(\beta(t)))$, where α (resp. β) are continuous non-decreasing functions from $[0, 1]$ to $[a, b]$ and d is a Euclidean distance between the points. Alt and Godau (1995) developed an algorithm for polygonal curves which finds exact Fréchet distance in $O(nmlog^2nm)$ time, where n and m are the number of segments on each curve. For a discrete version of the problem, in which curves can be aligned only on a finite number of points, an algorithm exists which computes the distance using dynamic programming in $O(nm)$ time. Fréchet distance was successfully used in different research fields for tasks such as handwriting recognition (Sriraghavendra, Karthik, & Bhattacharyya, 2007) or protein alignment (Jiang, Xu, & Zhu, 2008). Because the raw data from the eye tracker can be described as a discrete curve, we can use the discrete variant of Fréchet distance for measuring similarity of the curves. To our knowledge, this

[3]Generally, curves can be defined for any metric space; but for our purposes, restriction to Euclidean space is sufficient.

metric has not been used for scan pattern comparison with the exception of a bachelor's thesis by Kocián (2014).

2.4 Related work on metric comparison

As we have introduced above, there are many different metrics used for comparing scan patterns. Although there is some research on the experimental comparison of saliency map measures (Riche, Duvinage, Mancas, Gosselin, & Dutoit, 2013), there are only a few studies that include comparison of this metric with respect to scan pattern variability. Jarodzka et al. (2010) showed eight pairs of scan patterns in their work. Each of the patterns represented one scenario for how two scan patterns could differ. The scenario they used was as follows.

- Spatial offset – Scan patterns can differ in spatial offset, where one scan pattern is systematically shifted in one spatial dimension (by adding some constant value to one spatial coordinate).

- Temporal offset – Scan patterns can be identical with respect to spatial position, but the time coordinate can lag by a given constant.

- Reversed order – The spatial position can be identical for both scan patterns, but they would be visited in reversed order.

- AOI border problem – Fixations in one scan pattern are close to the AOI borders. The second scan pattern contains fixations that are in the neighboring AOI. Therefore, they are spatially very close to each other, but methods working with AOI would treat them very differently.

- Scaling – Two scan patterns are identical copies of each other, but the second one is a scaled version of the first one. Therefore, the shape and temporal coordinate are identical, but the spatial coordinates differ.

- Local/Global – This change keeps the overall similarity of the shape, but it changes long saccades into several consecutive small ones. This captures whether the inspection of the scene is local (small saccades) or global (long saccades).

- Duration – Two scan patterns are identical with respect to the spatial coordinates, but they differ in the duration of individual fixations.

They compared correlation of the saliency maps, Levenshtein distance and Multimatch. Multimatch showed the best results (at least one dimension showed similarity with respect to the given change). They did not, however, systematically vary the parameters for each transformation. Their work is important for showing how scan patterns can differ in the space of possible transformations.

Another study from Dewhurst et al. (2012) showed the difference in scan pattern comparison more systematically. In their first experiment, this group created pairs of random scan patterns and created a copy of one of the scan patterns by adding spatial Gaussian noise to each fixation. Then, they used both MultiMatch (measuring several similarity measures at once) and ScanMatch to test whether they could find the original scan pattern and a copy more similar than the other one. Despite the slight noisiness of the data, both ScanMatch and MultiMatch correctly identified that the modified and original version were more similar. With an exception for the scenario with high spatial noise, MultiMatch outperformed ScanMatch (in the dimension-measuring closeness of the aligned fixations). In their second experiment, they used scenarios similar to those in Jarodzka et al. (2010). They compared the similarity computed by MultiMatch and ScanMatch for each scenario; and also for two random scan patterns as well as a baseline. The results showed that MulitMatch outperformed ScanMatch for all operations: with exceptions for temporal offset, in which they scored similarly. In comparison to the random baseline, at least one dimension of Multimatch showed better results than the similarity of the random scan pattern.

Another comparison of ScanMatch and Levenshtein distance was performed by (Cristino et al., 2010). Similar to Dewhurst et al. (2012), they varied varied the added noise in the encoded strings. ScanMatch outperformed Levenshtein distance for all noise levels.

All of the above-mentioned comparisons included only methods that represented scan patterns as a succession of events. For representations using raw samples (and especially for scan patterns from dynamic tasks with a prevalence of smooth pursuit eye movements), there is only work from Kocián (2014). In his thesis,

he created artificial scan patterns using evolutionary algorithms to match behavioral data from the experiment presented in Děchtěrenko et al. (2017). Similar as Dewhurst et al. (2012), he created pairs of random scan patterns, and, to one of the each pair, he applied one of the following transformations: spatial offset, rotation, scaling and flipping the scan pattern around a horizontal axis. Three metrics were selected for comparison: Normalized Scanpath Saliency, Levenshtein distance and Fréchet distance. The Levenshtein distance accuracy decreased quickly for all transformations and scored worse. The Fréchet distance attained an almost perfect score for both spatial offset and scaling for all levels of transformation. The NSS attained a perfect score for spatial offset < 2 DVA and scaling coefficient < 1.4 (the modified version of the scan pattern was 1.4 times larger), and then accuracy gradually decreased. For the rotation, the Fréchet distance outperformed NSS for angles $< 100°$. As expected, all metrics scored poorly for the flipping of the axis.

2.5 Scalability of metrics for raw samples

Although there are many different metrics that can be used for scan pattern comparison, it is unclear how we can compare their values. Here, we try to vary the noisiness of the scan patterns and compare the similarity of the original and modified ones. We did four simulations in this section. First, we showed the relationship between the Normalized Scanpath Saliency (NSS) and Correlation distance (CD). Second, we tested the behavior of CD when comparing three spatio-temporal maps at the same time. Third, we selected scan patterns from the behavioral experiment (see section 3.5) and tested the robustness of four metrics when we applied three transformations similar to Dewhurst et al. (2012). Finally, we applied the transformations systematically on the scan patterns and computed their similarity to the original ones.

2.6 Simulation 1 – NSS versus Pearson correlation

In our previous experiments (Děchtěrenko & Lukavský, 2014; Lukavský, 2013), we used the NSS metric to compare scan patterns. As we will show in this section, this metric has two drawbacks. First, its maximum value is dependent on the number of saccades in the data. Second, it is harder to interpret the values because of their arbitrary scale. The similarity of scan patterns expressed as a Pearson correlation coefficient solves the issues. Therefore, we tested the relationship between the Pearson correlation and NSS in three scenarios.

2.6.1 Methods

Metrics

For all three scenarios, we used two metrics for scan pattern comparison: NSS and Pearson correlation. NSS was computed as described in Section 2.3.3 – one artificial scan pattern was convolved by a spatio-temporal Gaussian filter forming a spatio-temporal fixation map. Then the similarity of the second scan pattern was computed relative to this map. The Gaussian filter had the parameters $\sigma_x = 1.2°$, $\sigma_y = 1.2°$ and $\sigma_t = 26.25$ ms. However, as shown by Lukavský (2013), similar results were obtained for filters with different parameters. The Pearson correlation was computed as follows. First, each scan pattern was convolved with a spatio-temporal Gaussian filter (the same parameters as in the NSS case) and the Pearson correlation r coefficient was computed between the two maps. For easier computation, we binned each scan pattern into a spatio-temporal matrix ($0.25° \times 0.25° \times 20$ ms) before the convolution. We computed the similarity between the original scan pattern and the modified versions.

Scan patterns and modifications

For both scenarios, we created 50 artificial scan patterns as a random walk process, in which each subsequent position was sampled from bi-va-ri-a-te normal distribution centered at the last position with a covariance matrix $0.05 \cdot I$, where

parameter 0.05 was selected so that scan patterns showed visual resemblance to the real scan patterns. Each scan pattern consisted of 500 samples. The time coordinate in this case is arbitrary, but we treated the time coordinate as if it were generated with a sampling frequency of 250 Hz, so the total length of the scan pattern would be 2 s. An example of artificial scan pattern is shown in Figure 2.3. To establish the relationship between the metrics, we created three scenarios in this simulation. In the first scenario, we applied two transformations to the artificial scan patterns simultaneously:

- Spatial offset – we translated each scan pattern by 0°-4° (step size of 0.25°)
- Incoherence – we translated portion of the scan pattern by a large value (10°). This distance shows highly incoherent scan patterns. The portion of the scan pattern ranged from 0 (identity) to 1 (completely different scan patterns), with a step size of 0.1.

Therefore, we had 21 * 11 = 187 modifications for each scan pattern.
In the second scenario, we tested how NSS and the Pearson correlation scale when we add artificial saccades. This is relevant to real world application. For example, in the Multiple Object Tracking task, one of the strategies used for tracking involves switching between targets. Therefore, in the second scenario, we changed the number of times in which the scan pattern alternated between two positions 5° apart. The number of changes were varied from 0 (identical scan pattern) to 29 with a step size of 1. To capture the whole range of NSS values, we also added a higher number of switches, starting from 30 to 230 with a step size. Therefore, for the case with 1 shift, the scan pattern changes its position once by 5° and stays there. While for the case with 230 shifts, the scan pattern alternates 230 times between the shifted and non-shifted position.
For the second scenario, we had 36 modifications for each scan pattern.

2.6.2 Results

For the first scenario with spatial offset and incoherence, both NSS and the Pearson correlation coefficient behaved similarly for both transformations (Figure 2.4). We plotted contours showing the decrease in values for both metrics. The contour

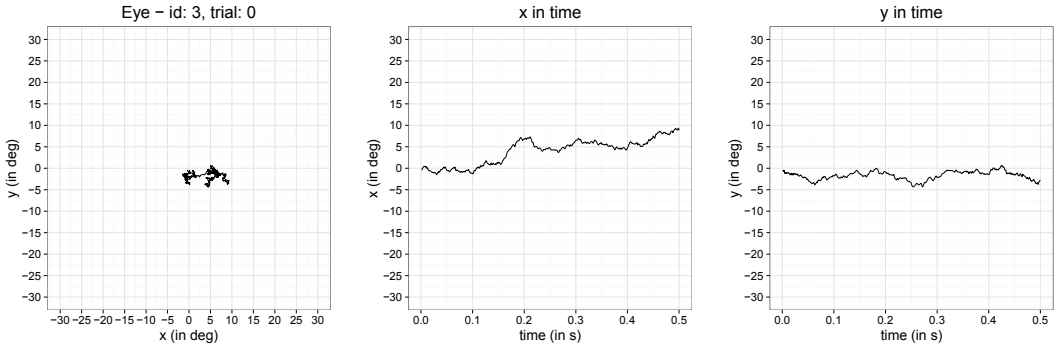


Figure 2.3: Example of an artificial scan pattern. The first plot shows the scan pattern in space; the last two plots show each coordinate in time.

denotes combinations of transformations with identical NSS or the Pearson correlation. The Pearson correlation decreased more gradually for the spatial offset. When one metric is plotted against another, the values are highly correlated (Figure 2.5, $r = .97$). The maximum value of NSS is dependent on the number of switches; it decreases rapidly from the maximum value for NSS (22.76 in this case) to the baseline.

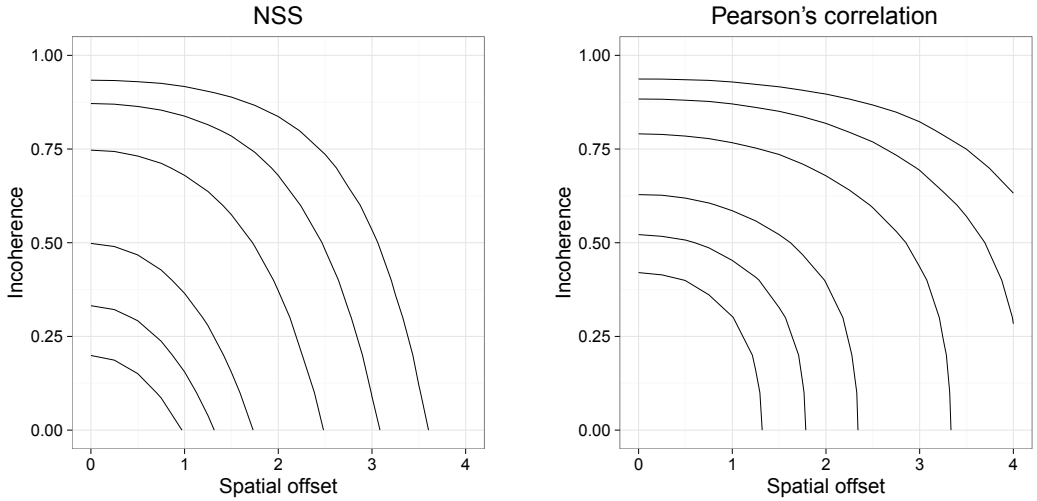


Figure 2.4: Normalized Scanpath Saliency and Correlation distance in the first scenario. The value shows the similarity between the original artificial scan pattern and the modified versions. Each contour shows identical levels of similarity for combinations of transformations.

In the second scenario, the similarity of two identical scan patterns varied from 22.55 for the scan pattern without any shift to 15.74 for scan patterns without

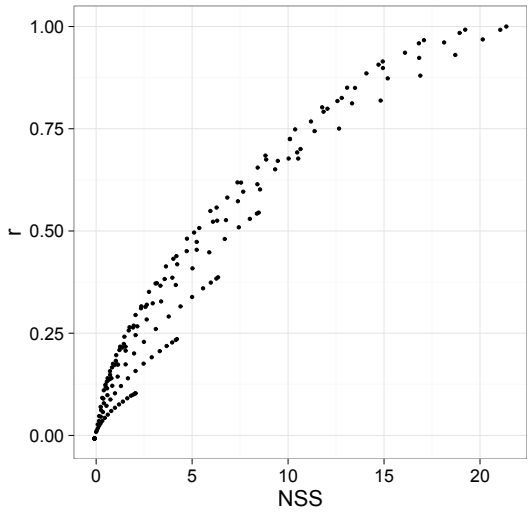


Figure 2.5: Correlation between NSS and the Pearson correlation. The values are highly correlated.

230 shifts (Figure 2.6). The Pearson correlation coefficient remained 1 for all scan patterns.

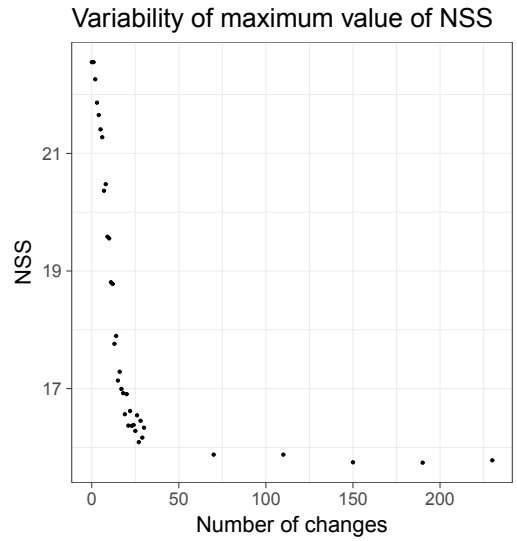


Figure 2.6: Maximum value of NSS dependent on number of switches.

2.6.3 Discussion

Results showed that both metrics performed similarly when we applied spatial offset and incoherence to the scan patterns. However, both values were highly correlated ($r = .97$). Therefore, we get a similar idea about similarity when

comparing the values across different experiments. An important drawback of NSS is the lack of a maximum value. The NSS value is related to the length of the trajectory and to the binning parameter, while the Pearson correlation has maximum distance. We did not test the effect of the binning parameter, because it was tested in Lukavský (2013).

2.7 Simulation 2 – Correlation distance

OOur previous simulation showed good resemblance of the Pearson correlation and NSS. However, the problem with the Pearson correlation is that it computes similarity, not distance. In this experiment, we created a measure of similarity called Correlation distance (CD) and tested its performance in the artificial scenario. In particular, we were interested in how the metric behaves for spatial offset and multiple scan patterns at once.

2.7.1 Methods

Metrics

We defined Correlation distance as CD = 1 – r, where r is the Pearson correlation between two spatio-temporal maps. Although in the case of saliency maps, the correlation coefficient could reach zero. For the spatio-temporal maps, it reaches values below zero only occasionally. Therefore, we set a negative correlation to 0 to get the maximum distance of CD equal to 1.

Scan patterns and modifications

Similar to the previous simulation, we created artificial trajectories as a random walk. Our scenario looked as depicted in Figure 2.7. We generated three trajectories; the initial position is denoted by red dots. Black dots show the starting position of modified scan patterns. In this setting, we tested three scenarios.

- 2 scan patterns – In this scenario, we created two spatio-temporal maps. The first one was created from the unmodified scan patterns 1 and 3. The second was created from the unmodified scan pattern 1 and from a modified

version of scan pattern 3 that was moved to the right (therefore the x-coordinate ranged from 5° to 10°).

- 3 scan patterns (one was translated) – Again, two spatio-temporal maps were created; the first one from the unmodified version of all three scan patterns and the second one from the unmodified version of scan patterns 1 and 2 and a modified version of scan pattern 3 (similar to the previous case).

- 3 scan patterns (two were translated) – This is identical to the previous scenario. However, in this case, scan pattern 1 moves to the left as well (therefore the x-coordinate ranged from −10° to −5°).

We had 50 repetitions of each scenario.

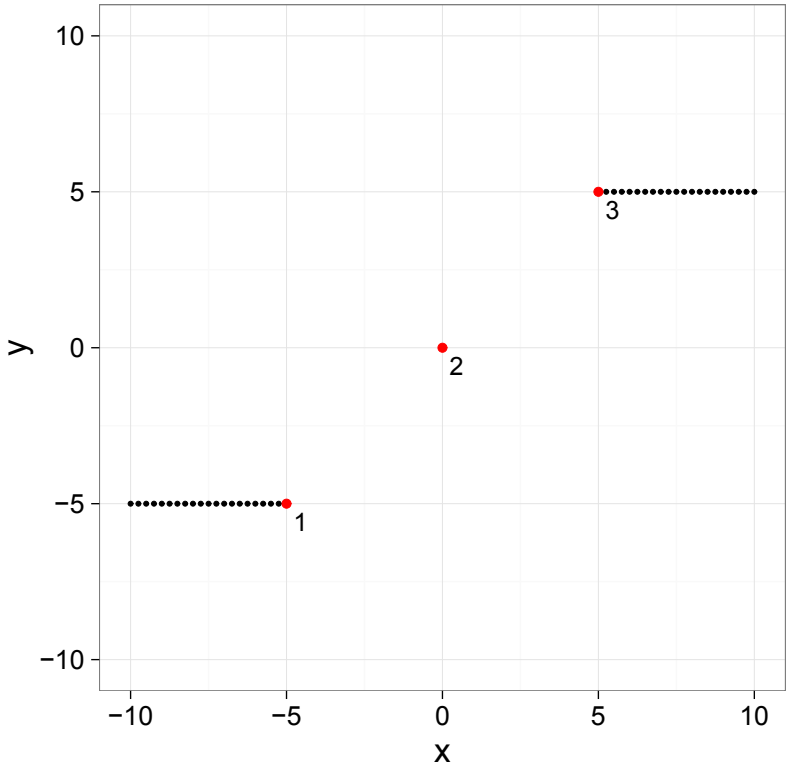


Figure 2.7: Initial position for generation of artificial scan patterns. Red dots denote the baseline, black dots denote scan patterns with added spatial offset (the number in the bracket is the scan pattern identifier). See text for detailed description of scenarios.

2.7.2 Results

As visualized in Figure 2.8, the correlation distance increases with the increasing spatial offset. It does so in an S-shaped pattern. For the case with two scan patterns (left), the distance reaches .5, for three scan patterns (one moves) it reaches .33, and for the three scan patterns (two moves) it reaches .66. Therefore, the metric behaved as expected.

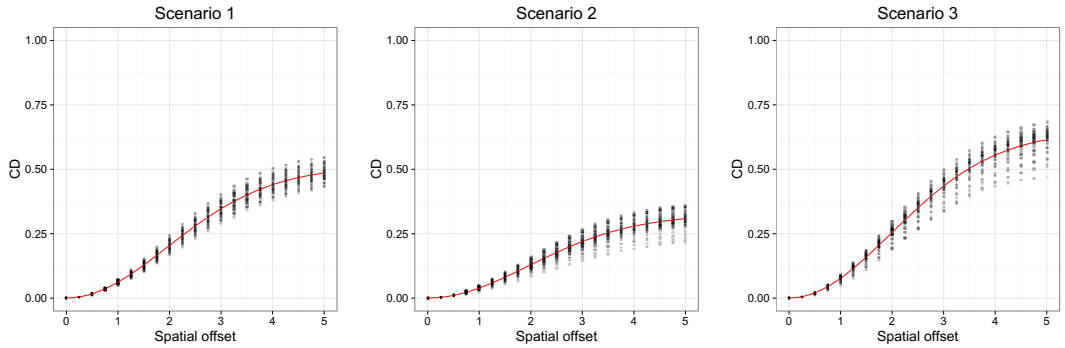


Figure 2.8: Average CD values for different spatial offsets. In each scenario, the distance increases in an S-shaped pattern. For the case with two scan patterns (left), the distance reaches .5, for three scan patterns (one moves) it reaches .33, and for three scan patterns (two moves) it reaches .66.

2.7.3 Discussion

The Correlation distance behaves as expected. For the case, where one half of the scan patterns was identical, the CD values reached .5 when the spatial offset for the second scan pattern was large. For the case with three scan patterns, it reached either .66 or .33 in dependence on whether two scan patterns moved or one. Taken together with the previous simulation, we can conclude that CD is useful metric for scan pattern comparison. Finally, this metric has clearly defined range and it behaves as a distance measure.

In the next section we will explore the relationship between CD and other methods that can be used for scan patterns represented using raw samples.

2.8 Comparison of the metrics

In the remaining part of this chapter, we try to establish a relationship between several metrics that can be used for comparison of scan patterns represented using raw samples. The CD metric showed promising results in the simulations and, therefore, we selected four more metrics for comparison to one another. Each metric was used to compute scan pattern similarity and its modification.
We selected the following metrics for scan pattern comparison:

- Mean and Median distance
- Levenshtein distance
- Correlation distance
- Fréchet distance

Mean and median distances were computed as the mean (median) of all Euclidean distances between temporally-aligned samples of scan patterns. This value corresponds to the average distance between two scan patterns. Other metrics were introduced in Sections 2.2 and 2.3. We did not aim for an exhaustive list of comparison metrics, but rather we wanted to relate the values of the CD metric to other metrics. Our design could be easily extended to other metrics as well. The Levenshtein metric was included because of its frequent use in literature. Mean and median distance of scan patterns gives intuitive results of the distance between the scan patterns. Finally, we included Fréchet distance, because of its promising preliminary results shown by Kocián (2014).
We used three transformations of scan patterns: translation, rotation and scaling. We systematically applied the transformation to the scan patterns and computed the distance between the original scan pattern and the modified one. We selected those three transformations since they correspond to situations that could happen during eye tracking sessions. Translation corresponds to a situation where viewpoint (e.g. the chin rest) is moved to either side. Rotation corresponds to situations where participants tilt the head sideways and scaling corresponds to situations where viewpoint is closer to/farther from the display. In addition, for the MOT task, those operations would be easily applicable to object trajectories.

For example, we could rotate the trajectories of all moving objects by 30°.
We decided to use behavioral scan patterns instead of artificial ones. To obtain a general result, it would be better to use larger groups of scan patterns for evaluating the metrics. However, the simulations were computationally demanding, so using the group with many scan patterns would have resulted in long simulation runs. Therefore, we decided to study variability on only one the CD metrics first and test how the obtained average CD values would differ when we use different subsamples.
Originally, we planned to use ScanMatch or other event based metrics for the comparison between the scan pattern and its modification. Although it should be possible to classify scan patterns using modern methods that are able to detect smooth pursuit, there is a problem with how to encode smooth pursuit into the grid. The length of fixations could be easily encoded using additional letters, but similar encoding is not possible for smooth pursuit. We could encode the start and end points for smooth pursuit using special characters, but this would omit the shape of smooth pursuit. Another possibility would be to encode each position using a special character, but this approach would reduce the data to the raw-sample characterization of scan patterns. Therefore, we decided not to include this metric in the analysis and focus on the above-mentioned metrics. Similar problems would also arise for the other metrics.

2.9 Simulation 3 – Variability of CD metric in dependence on number scan patterns

2.9.1 Methods

For each scan pattern, we applied each of the operations individually (we did not apply combinations of operations on each scan pattern, due to time complexity).

Metrics

We used a correlation distance similar to the previous Simulation 2. Variability of CD values was expressed as the size of the confidence interval of all correlation

distances for each subset and each operation.

Scan patterns and modifications

We randomly selected 160 distinct scan patterns from the MOT experiment (described in Section 3.5). Each scan pattern corresponded to 6 s long tracking period and consisted mostly of smooth pursuit. Then we randomly selected subsets repeatedly from the preselected scan patterns ($N = 40$) and varied the number of scan patterns in each subset (n = 10, 20, 40, and 80). ($n = 10$, 20, 40, and 80). From each scan pattern, we created a modified copy by applying one of the following three transformations:

- Translation – scan patterns are translated by 0–4 DVA[4] (step size of 0.25 DVA)
- Rotation – scan patterns are rotated by -15–$15°$ (step size of $2°$)
- Scaling – scan patterns are scaled by factor 0.5–1 (step size of 0.1)

Examples of each transformation are shown in Figure 2.9. Red color denotes the original scan pattern; blue color denotes a modified scan pattern. We applied each transformation individually. To avoid ambiguity, we used the abbreviation DVA (standing for degrees of visual angle) instead of the symbol ° that is used as a unit for rotation transformation.
Variability of correlation distance was measured as the size of the confidence interval of all correlation distances for each subset and each operation: denoted here as sCI.

2.9.2 Results

The size of confidence intervals (sCI) increased for scaling and rotation; for the translation, the increase was small. For the rotation, it reached values of 0.5–0.75 (size of 1 would correspond to a situation, in which at least one sample from the 160 trajectories of size n would have zero distance and at least one sample would have maximum distance of 1). The shape of the relationship was identical for

[4]We remind that DVA stands for degrees of visual angle

Figure 2.9: Examples of all three transformations. Scan pattern is translated by 3 DVA (left), rotated by 30° (middle) and scaled by .7 (right)

groups of all sizes, with different intercepts for each group. The average size of confidence intervals for all steps together is shown in Figure 2.11. For the scaling, the shape was similar (the scaling factor has inverted scale, so the curve is reversed as well). For the translation, the increase in the sizes of the confidence interval was small. This is not surprising because, although spatial offset increases mean distance, it does not increase overall variability for the groups of different sizes.

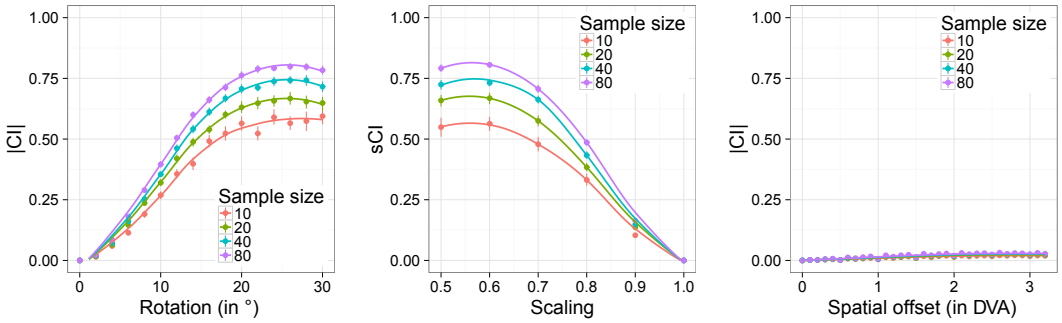


Figure 2.10: Size of confidence intervals for correlation distance. Correlation distance is computed for each step.

2.9.3 Discussion

Based on our results, we decided to select 20 scan patterns as a compromise between the variability of scan patterns and time needed to finish the computation. The difference between the mean value of sizes of confidence intervals for samples with 20 scan patterns and samples with 80 scan patterns was 18% for the rotation and 18% for the scaling as well.

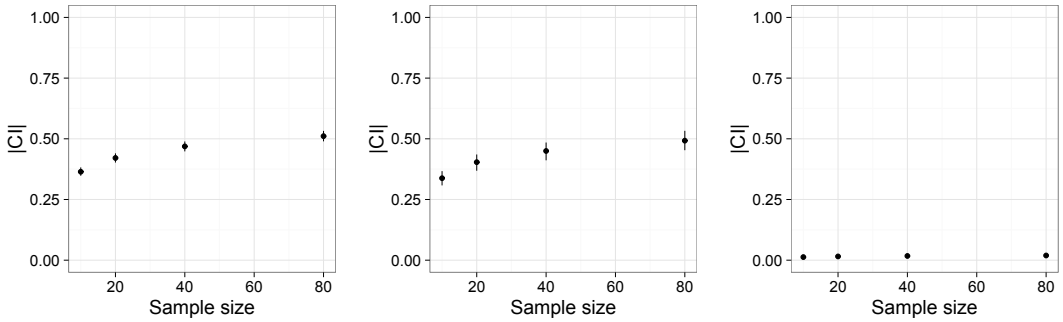


Figure 2.11: Size of confidence intervals for correlation distance for all sample sizes together. There is a similar pattern for rotation (a) and scaling (b). Translation does not increase the size of confidence intervals (c).

2.10 Simulation 4 – Robustness of the metrics

In this simulation, we applied the same three transformations as in Simulation 3 (translation, rotation and scaling) to the scan patterns from the MOT task. Then we selected five metrics and evaluated the similarity between the original scan pattern and the modified one. We also present tables that show the values of the selected five metrics dependent on the applied transformation, and we include tables that show how the transformations are related to each other in terms of scan pattern similarity.

2.10.1 Methods

Metrics

We used Mean and Median distance, Levensthein distance, Fréchet distance and Correlation distance to measure similarity. For Mean and Median distance, we used Euclidean distance as a measure of distance between temporally-aligned samples of scan patterns. For the computation of Levenshtein distance, we divided the display into $2.5° \times 2.5°$ rectangular AOIs. Because each scan pattern varied from $-15°$ to $+15°$, samples of the scan patterns could belong to one of the 169 possible AOIs (13 * 13). We decoded raw samples in each scan pattern using the AOI code and forming a large string. The distance between strings was computed using the Levenshtein algorithm. Finally, string distance is normal-

ized to a range 0–1. Regarding Fréchet distance, we first computed Euclidean distance between all samples of both scan patterns. To obtain the pairing with minimum distance, we used dynamic programming similar to what we did with Levenshtein distance. Correlation distance was used similar to previous simulations. Each scan pattern was compared to all of the modifications. We used the following metrics for evaluating distance between scan patterns. Each scan pattern was compared to all of the modifications. We used following metrics for evaluating distance between scan patterns.

Scan patterns and modifications

We selected 20 distinct scan patterns from the Multiple Object Tracking experiment. For each scan pattern, we applied three transformations: translation, rotation and scaling as we did in Simulation 3. In this case, we applied two operations transformations simultaneously. Applying all three operations at the same time would result in a large number of calculations per each scan pattern (41 * 16 * 5 = 3936 modifications per each scan pattern). If we imagine the operations as three orthogonal axes, then we are working with individual planes instead of working with a whole cube. We had 656 modifications for the Translation and Rotation operations (fixed Scaling to 1), 246 modifications for Translation and Scaling operations (fixed Rotation to 0) and 96 modifications for Rotation and Scaling operations (fixed Translation to 0).

2.10.2 Results

As denoted in Table 2.1, correlation between metrics is high. The lowest correlation was between Fréchet distance and other metrics (r = .70–.85). All correlations were significant with $p < .001$.

The application of each transformation separately is visualized in Figure 2.12 (page 66). Each row shows the average values for each metric. We can see that the CD metric had the same S-shaped pattern as in our other simulations. For increasing transformations, the Levenshtein distance increased nonlinearly. For the translation, the largest standard error for the mean was for values around 1 DVA. With the exception of Levenshtein distance, all other metrics showed no

	CD	Fréchet	Levenshtein	Mean
Fréchet	.79	–		
Levenshtein	.91	.70	–	
Mean	.97	.85	.85	–
Median	.96	.82	.84	.99

Table 2.1: Correlation between values for each individual metric.

variability for the spatial transformation. This is not surprising, because adding spatial offset to one coordinate is not related to the shape of the scan pattern. Therefore, there is no variability. In the case of Levenshtein distance, the source of variability for the spatial transformation is due to the problem with AOI borders (small differences in spatial noise could result in falling into different AOIs). Exact values that can be used for comparison between metrics can be found in Appendix A in Tables A.1, A.2, and A.3.

For cases where two transformations were applied at the same time, the values of each metric is shown in the Figures 2.13, 2.14 and 2.15. For the combinations of transformations, all metrics behaved similarly. However, Fréchet distance decreases faster than other metrics. The Figures show contours for different levels of each operation. The upper row of plots shows contours for fixed levels of transformation on the x scale; the lower row of plots shows contours for the fixed transformation on the y scale. For example, in Figure 2.13, the first plot in the upper row shows four contours that correspond to the scan patterns that are only rotated by values of 5, 10, 15 and 20°. Exact values showing how the scan pattern is distorted by one transformation related to a distortion in another dimension is shown in Tables A.4, A.5, and A.6.

Note that those transformations are small in comparison to the random baseline. We computed the distances between all pairs of original scan patterns using each metric and average values used. Because the scan patterns represented the tracking patterns for the random moving objects, they are not spatially aligned. To correct that, we align the scan patterns, so that they would only differ in their variability and shape. Table 2.2 shows mean and SD values (in brackets) for each metric and each type of alignment. Overall average denotes the alignment of scan

patterns, i.e. that both x- and y- coordinates do not differ from zero in average. The midpoint denotes alignment in which the midpoint of both scan patterns has coordinates equal to zero. The first sample denotes alignment, where the first sample of the scan pattern has coordinates equal to zero. The results show that, even after alignments, the distance between two scan patterns is still large. For CD and Levenshtein, the distance between two random scan patterns is around 3 DVA. The average distance between random scan patterns evaluated by the remaining three metrics is larger than the distances between any transformed scan pattern and its original.

Metric	Overall average	Midpoint	First sample	original
CD	0.84 (0.14)	0.80 (0.10)	0.91 (0.08)	0.94 (0.07)
Fréchet	9.39 (3.71)	10.69 (4.13)	10.78 (4.34)	11.06 (3.51)
Levenshtein	0.90 (0.12)	0.86 (0.07)	0.94 (0.05)	0.98 (0.03)
Mean	5.20 (1.88)	5.76 (2.16)	8.15 (3.22)	7.33 (2.23)
Median	5.08 (2.06)	5.84 (2.49)	8.01 (3.46)	7.00 (2.50)

Table 2.2: Distance between two random scan patterns. Scan patterns were aligned to the central point by several methods (the overall average is zero, the midpoint sample is zero, the first sample is zero) or the original locations were used.

2.10.3 Discussion

In this simulation, we studied the behavior of several metrics for scan pattern comparison (Correlation distance, Fréchet distance, Levenshtein distance, Mean and Median distance) when the scan pattern is modified by one of three transformations (translation, rotation and scaling). We selected those three transformations for two reasons. First, because they correspond to the possible source of error during eye tracking experiments. Second, because in MOT, object trajectories can be easily and systematically modified. Those operations on scan patterns correspond to the simple transformation of object trajectories in MOT. There are more possible transformations that could be applied to the scan patterns. Therefore, our code could be easily extended for other transformations and other

metrics. Out of the three selected transformations, translation is the most useful one. It corresponds to the average spatial offset of two scan patterns. Therefore, it describes the situation even with distinct scan patterns and not just the spatial transformation of the identical scan pattern (as in our case). All of the metrics decreased in a similar way.

We introduced a novel method for comparison of scan patterns in the context of eye tracking – Fréchet distance. This method was used for estimating the distance between general curves (Jiang et al., 2008; Sriraghavendra et al., 2007). As our representation of scan patterns shows high resemblance to general time series, we can apply more methods from different research areas. Fréchet distance is an example of such an approach.

We also showed the baseline for distance between random scan patterns. We used those values similarly to Dewhurst et al. (2012) or Jarodzka et al. (2010) to test discriminability between random scan patterns and modification of scan patterns. However, the distance between two random scan patterns was larger than any distance between the original and modified scan pattern for Fréchet, Mean and Median distance. For CD and Levensthein, scan patterns were closer in the random scan pattern than in the modification only for spatial offsets larger than 3 DVA. This replicates the findings of Kocián (2014), in which Fréchet distance showed perfect discrimination for transformed scan patterns.

Our results can be used in two ways. First, they show the relationship between individual metrics, so we can translate our results from one study to another. Second, we can translate the results of our study in terms of scan pattern variability. This means that we could achieve intuitive understanding of the variability of scan patterns per each condition or even the difference between groups. In the following chapters, we use this approach to explain the differences between experimental conditions.

Our approach can be applied in other research areas as well. Irrespective of the type of time series, the evaluation of metrics for comparing similarity is important. A similar approach could be used for establishing the relationship of the metrics in different contexts. This would simplify interpretation of the results of different studies in the context of other studies.

2.10.4 Limitations

Our analysis has several limitations. First, we only used three transformations of the scan patterns. There are other possibilities for altering the scan patterns. Therefore, to obtain more robust results, additional transformation should be included. This would require massive computational power which was not available to us.

Second, we selected only a few metrics. It would be interesting to extend the comparison to additional metrics used for scan pattern comparison such as Kullback-Leibler Divergence or the Percentile metric. In addition, the exact values for our metrics are related to the parameters of the metrics that we used. For example, values for Levensthein distance are dependent on the number of AOIs; CD is related to the parameters of the Gaussian filter used for convolution. Therefore, when different parameters are used, researchers could follow our simulations and compute values for their chosen metrics and transformations.

Finally, to extend our results to a general case, we would need to compute the averaged values for each transformation step and each metric for larger samples ($N > 1000$). Our results could be biased, but because the overall variability increased approximately and logarithmically with respect to the sample size, the potential bias should be small.

For other research areas, the transferability of results is limited. Scan patterns are special types of time series – they contain discontinuities introduced by the saccades. Simulations 1 and 2 did not include the saccades, but Simulations 3 and 4 used real scan patterns. Therefore when applying our results for different time series, one should first replicate our comparison for time series that were measured in their research.

2.11 General discussion

In this chapter, we explored the relationship between several metrics. In Simulations 1 and 2, we explored the relationship between NSS and the Pearson correlation on the artificial scan patterns and established a Correlation distance metric. Although this metric has been used in several studies so far (Děchtěrenko

et al., 2017; Lukavský & Děchtěrenko, 2016), the properties of this metric had not been previously explored. Thus, we showed the relationship between this metric and other metrics that are typically used for scan pattern comparison. In the last two Simulations 3 and 4, we showed the behavior of the metrics when we systematically modified real scan patterns. The shape of the relationship between the transformations looked similar for all metrics. For the Experiments 1–4 discussed further in this thesis, we decided to use Correlation distance as a measure of distance between two scan patterns. The exact values of the CD metric for each transformation will be used to get a general sense of variability. We could look at the tables as an estimate of effect size for differences between scan patterns.

Another conclusion from Simulation 4 would be the introduction of Fréchet distance as a measure of scan pattern similarity. During the application of two transformations at the same time, the Fréchet distance behaved differently than the other metrics; especially for the case where spatial offset was added to the scan patterns.

Although we worked with special types of time series, our approach could be extended to other research fields. In particular, similar estimations of relationships between metrics could be used for other time series. For Simulations 1 and 2, time series could be generated with different parameters to show their resemblance to the time series in the research. Then, for Simulations 3 and 4, time series that were used in the research should be used as well. Other research fields could use metrics common in eye tracking research. In particular, CD and NSS could be easily transfered; the only problem would be the estimation of parameters for the Gaussian filter that we used for convolution. Our parameters were based on parameters for eye movement data. However, as shown by Lukavský (2013), the metrics do not differ qualitatively when different parameters are used. Therefore, with slight adjustments, the metrics could be used in different contexts as well.

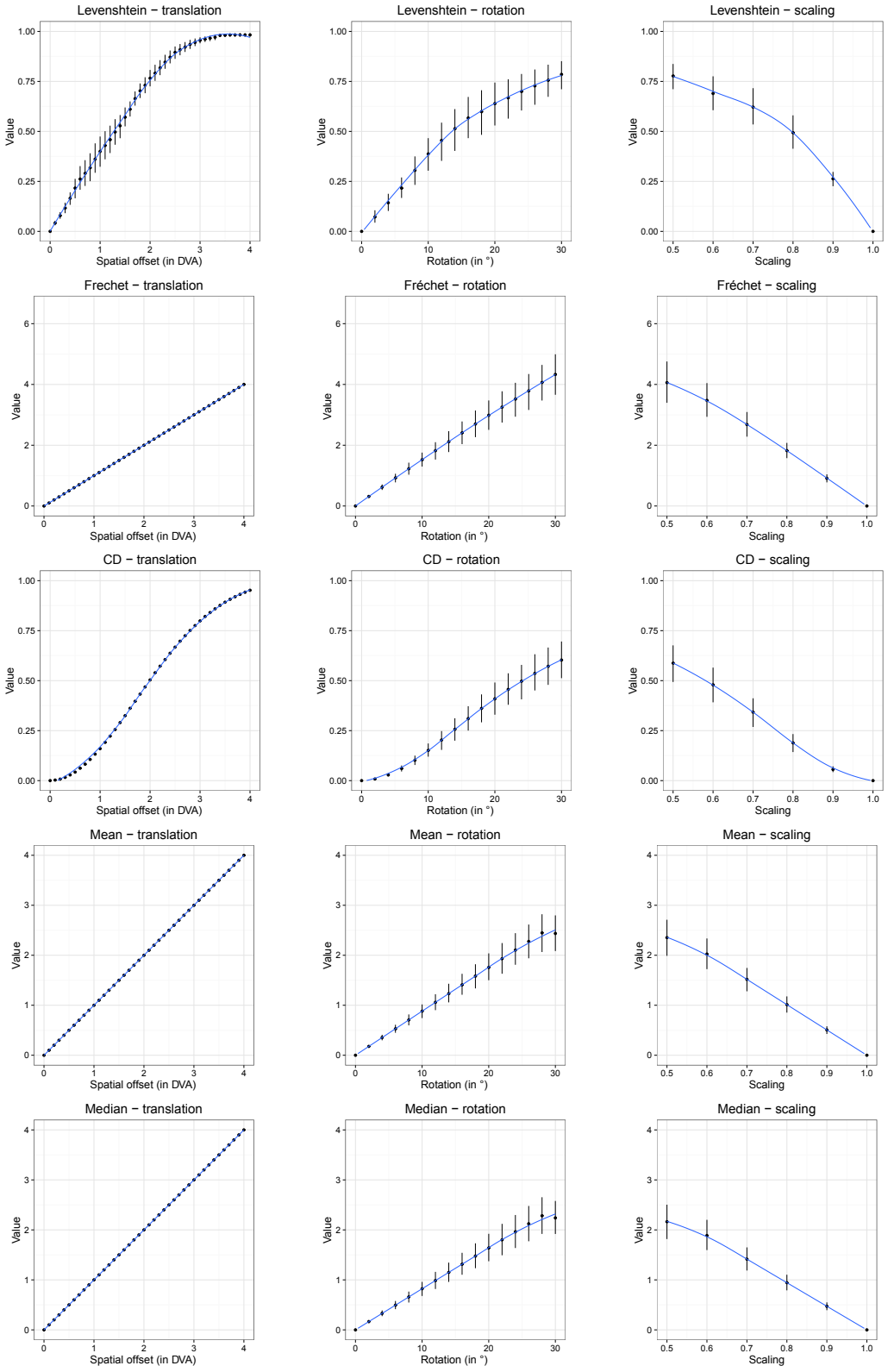


Figure 2.12: Average values for each distance and each transformation. Black lines denote standard error of the mean.

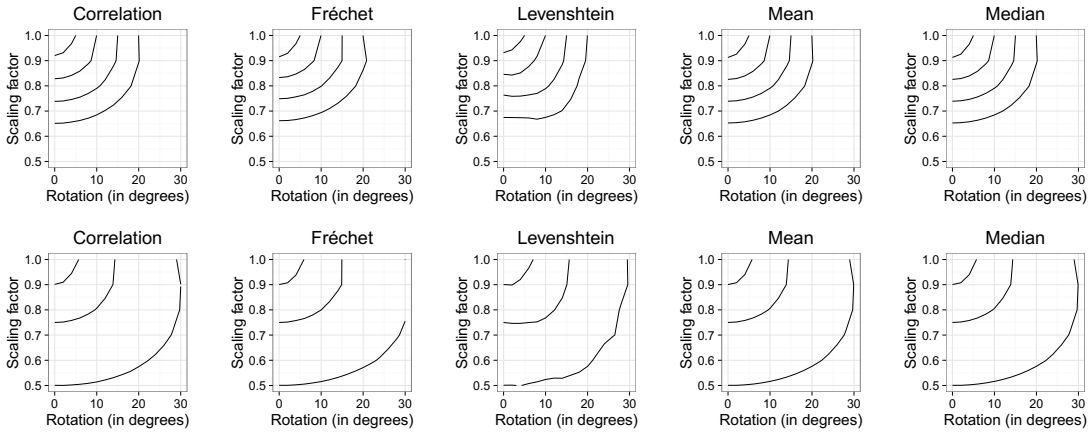


Figure 2.13: Values of the metric based on the combination of scaling and rotation. Lines denote same values. Upper plot shows contours for fixed rotation (5, 10, 15, and 20°), ; lower for the fixed scale factor (0.9, 0.75, and 0.5).

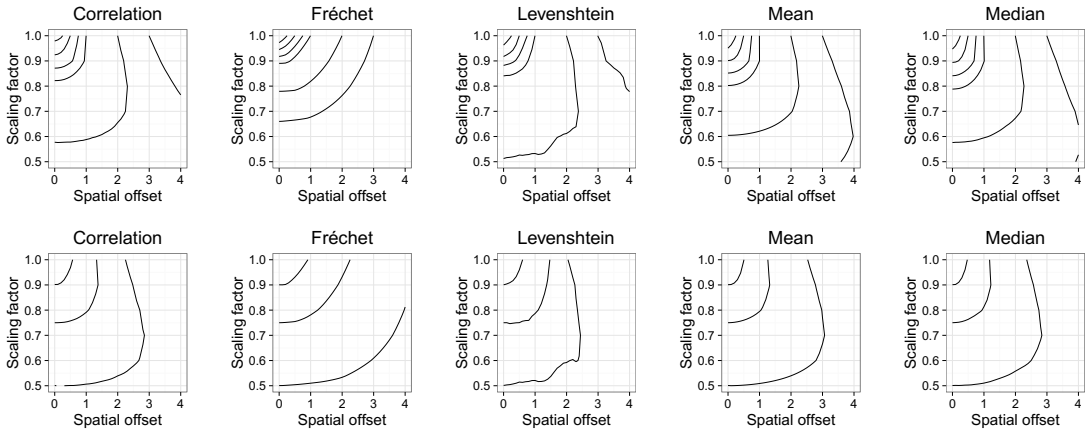


Figure 2.14: Values of the metric based on the combination of scaling and translation. Lines denote same values. Upper plot shows contours for fixed spatial offset (0.25, 0.5, 0.75, 1, 2, and 3 DVA); lower for the fixed scale factor (0.9, 0.75, and 0.5).

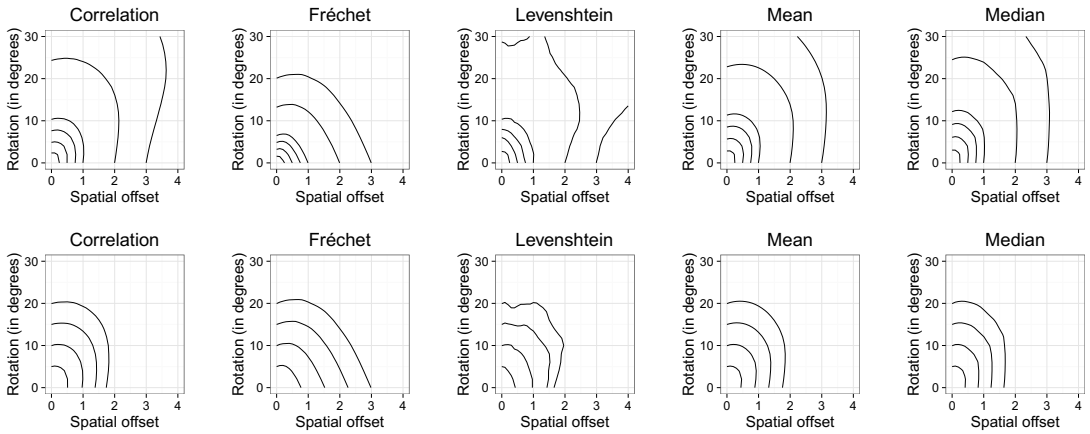


Figure 2.15: Values of the metric for the combination of rotation and translation. Lines denote same values. Upper plot shows contours for fixed spatial offset (0.25, 0.5, 0.75, 1, 2, and 3 DVA); lower for the fixed values for rotation (5, 10, 15, and 20°)

3. Significance testing for groups of scan patterns

Imagine the following scenario. We have two groups of participants (e.g. patients and healthy subjects) to whom we present the same movie. Did both groups look at the movies differently? If we are interested in the dwell times on specific objects (e.g. did patients look more at the faces than at the healthy subjects), we can set dynamic AOIs and compare dwell times for each group by standard tests. In the case of scan patterns, the answer is more difficult. We can compute pairwise distance between individual scan patterns, but it is unclear how we should test the differences statistically. In other words, we must discern whether scan pattern variability in each group is different than the variability between groups.

This is a problem that is conceptually different from testing whether scan patterns in one group are more similar to each other than scan patterns in other group(s). This problem can be solved by computing group coherence instead of distances between all pairs of scan patterns. For each group of scan patterns, we get one value capturing the coherence of scan patterns, and we can test similarity in each group using classical approaches such as linear (mixed) models.

This type of comparison is a general problem outside the context of eye tracking research. When we have two groups of time series (e.g. fMRI data for one brain area for patients and controls). We could take an interest in whether the time series differs between the groups. When the variability of a time series is low, we can test the differences using linear models with a dummy variable corresponding to belonging to the group. Here, we present an alternative approach.

Solving the problem of how to test differences between groups of scan patterns can help us use movies as stimuli and test complex hypotheses. Movies are very complex stimuli. There is a lot of visual, auditory and semantic content (even in a short clip) that could attract a participant's attention. Movies are also hard to parametrize, and it is hard to connect eye gaze location (when watching a movie) to individual features of the frame. For conceptual questions about scan patterns in dynamic scenes, Multiple Object Tracking is a feasible alternative.

It has two main advantages. First, participants need to sustain their attention during the entire trial, because if the target identity is lost during the tracking, it is impossible to retrieve it later. The number of stimuli on the scene is lower than in movies. Therefore, we can easily parametrize the task by modifying the moving objects. Second, it is possible to present identical trials repeatedly. Because people fail to recognize repetition (Ogawa et al., 2009; Lukavský, 2013), we can study the intra-subject variability of scan patterns as well. This is more difficult when using movies. When presenting the same movie repeatedly, scan patterns would differ to a large extent, because participants would concentrate on different features in the movies.

Although people fail to recognize repetition in identical trials, some filler trials need to be added between the repeated instances of the same trial to mask the repetition. This would lead to longer experimental procedures. Although it is feasible to create experiments lasting 60 minutes, results would be impacted by participants‘ increased tiredness. Thus, researchers could not test complex hypotheses. One way to reduce the chance of noticing the repetition would be to use geometrical transformation from the original trial.

3.1 Chapter description

In this chapter, we describe approaches from the literature for comparing scan patterns between groups. We introduce two new methods we developed and verify them in an experiment. Also, we show a practical approach to group comparison in a MOT task, in which we explore one of the possible geometrical transformations of the MOT trial and compare scan patterns from the original and transformed trials.

The main results of this chapter (Experiments 2–4) have been presented in the author's paper (Děchtěrenko et al., 2017). Here, we include the results and aim to provide broader explanatory comments.

We begin the section with an experiment that shows the simple case of comparing overall coherence between groups[1]. For that, we employed a MOT trial to answer

[1]The problem of testing differences groups between groups is present in both within-subject and between-subject designs. Here, we show the overall differences in the within-subject vari-

questions about eye data quality.

3.2 Experiment 1 – Effect of wearing glasses

The simplest case of group comparison is to compare difference in coherence between groups. In this case, we take one value for each group and compare differences using standard methods. In this section, we show one example of this comparison in an experiment testing the effects of wearing glasses during an eye tracking experiment. This experiment shows another application of a MOT task different from the conventional use for studying divided attention.
In this experiment, we presented MOT trials repeatedly while participants put their glasses on and took them off during the time course of the experiment.

3.2.1 Introduction

As stated in Section 1.4.5, quality of eye tracking measurements is a heavily studied problem. There are many models of eye trackers from different manufacturers and each of them may produce data of different quality. Nyström et al. (2013) showed a non-significant effect of wearing glasses on eye tracking measurement. However, in their study, they selected participants based on the condition of whether they wore glasses or not and compared the differences in accuracy between those groups. In our study, we decided to test the differences in within-subject design. In particular, we selected participants with myopia and measured the eye movements in a MOT task while they repeatedly took put their glasses on and took them off during the time course of the experiment. This experiment reflects current malpractice by eye tracker operators. When the participant is wearing glasses, it is hard to calibrate. The participants are sometimes asked to remove their glasses, and then they are calibrated without them. This is sometimes justified by the unsupported claim that eye tracker data would not differ much.
In this experiment, we tested the differences in scan pattern coherence when the subject was wearing glasses and when he/she was not.

ability.

3.2.2 Method

Participants

Thirty-six subjects (8 males; ages 19–27, mean 21.94) with myopia participated in the experiment. All of them were tested for color blindness using a Ishihara color blindness chart, and all of them had myopia. The self-reported number of diopters ranged from +0.5 D to +6.375 D (mean = 3.03; $SD = 1.64$). When two eyes differed in their number of diopters, a mean value was used. Originally, 38 subjects were recruited but two were excluded due to eye tracking error.

Apparatus and stimuli

The experiment was programmed in MATLAB with the Psychophysics Toolbox extension (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). The experiment was presented on a 19-in. CRT monitor with a resolution of 1024 × 768 and an 85 Hz refresh rate. Participants' heads were restrained with a chinrest at a distance of 50 cm from the screen. Gaze position was recorded using Eyelink II (SR Research, Canada) with a sampling rate of 250 Hz. A nine-point calibration procedure was used as the calibration procedure before the beginning of each block. Calibration was repeated several times, until "Good" accuracy was achieved for at least one eye (maximum error $< 1.0°$ and average error $< 1.5°$). Only the eye with the lower calibration error was selected for tracking. Drift correction was performed before each trial.

Stimuli used in the experiment consisted of eight gray circles (1°in diameter, RGB: [128, 128, 128]) moving on a black background (RGB: [0, 0, 0]). The circles moved at a constant speed of 5°/s. In each frame, the direction of movement for each dot was sampled from a von Mises distribution with the parameters: direction $\mu = 0$ and concentration parameter $\kappa = 40$. A von Mises distribution is a unimodal continuous distribution with support $(0, 2\pi)$, and therefore it is often used for generating direction. Parameter μ is a measure of location, and it controls where the central direction is facing (oriented). Parameter κ controls how closely the samples are concentrated around the central direction (it is inversely proportional to the variance). For $\kappa = 0$ the distribution is uniform, for large values (such

as 40 in our case) the distribution approaches normal distribution with a mean of μ and variance $1/\kappa$ (Mardia & Jupp, 2000). Sampling from this distribution resulted in Brownian-like motion of the objects. All objects moved in a circular arena (30° in a diameter) and bounced off the invisible boundary back to the central area. The objects did not follow the laws of reflection, because that would have resulted in predictable direction changes (bouncing off the circular arena would result in copying the shape of the polygon). Instead, they bounced back to the central area with a random change in direction sampled uniformly from the interval $(-\pi/2, \pi/2)$. Objects also bounced off an invisible envelope surrounding each of them and allowing for at least 0.1° of space between them. When bouncing off each other, they objects obeyed the laws of reflection.

Procedure

Before the experiment, 18 participants were tested for visual acuity using the Freiburg Visual Acuity Test (Bach, 2007). They were tested with both their glasses on and off. The score from this test with glasses was denoted as FAcT-glasses (respectively FAcT-noglasses). The remaining 14 participants were asked for their number of diopters only.

There were 5 blocks in the experiment. The first 4 blocks were divided into three parts: one calibration block and 2 microblocks; the last block had only a calibration block and one microblock. Participants were without glasses during blocks 1, 3 and 5; and with glasses during blocks 2 and 4. Eye tracker was calibrated in each calibration block, and 2 unique trials were presented after the calibration procedure so participants could get used to the change in acuity. A nine-point calibration procedure was used. During the calibration, we calibrated both eyes, and then selected the eye with the higher accuracy. A few times during the calibration where glasses were left on, the reflection from the glasses forced us to select one eye manually prior to the calibration.

There were 72 trials divided into 9 microblocks (8 trials in each microblock). Three trials were identical to the trials from the previous block; three trials were unique and were presented in the next block; and the remaining two trials were unique and were not presented again (they served for masking purposes only).

The order of the trials in the microblock was randomized, but we ensured that no two repetitions of the same trial were presented in a row. In the first and last microblock, there were only 3 experimental trials; the other 5 were presented for masking purposes only. A detailed experimental scheme is depicted in Figure 3.1. In each trial, 8 objects were placed randomly within the arena; four of them were green (targets, RGB: [0, 255, 0]), while the other four remained gray (distractors). After 2 seconds, the targets changed their color to gray so as to become indistinguishable. Then all objects moved for 6 s. The participants' task was to track the original targets during movement. After 6 s, all objects stopped and participants selected all four targets. After selection, the participant was informed whether he/she selected all four targets correctly (green word "OK") or how many objects were incorrectly selected (red number). Selected objects changed their color to yellow (RGB: [255, 255,0]).

To reduce the chance of noticing the repetition, we generated an 8 s period of motion and then started randomly at some time point between 0 s and 2 s. Since each trial lasted 6 s, we had 4 s of common movement that all trials shared in the worst cases of overlap. Th whole experiment lasted approximately 45 min (including the positioning of the eye tracker and the calibration procedure). There

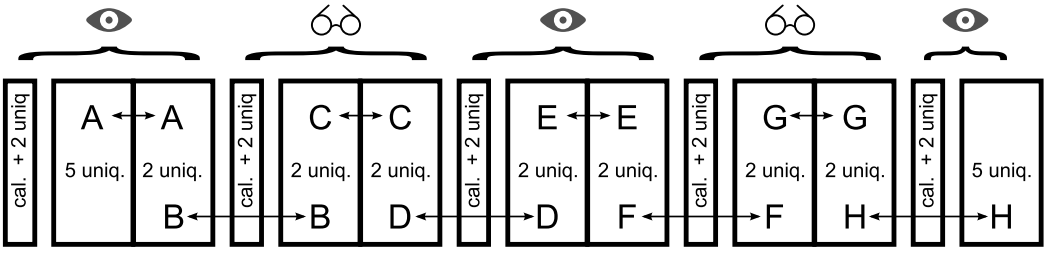


Figure 3.1: Experimental design. Trials were divided into 5 blocks; the first 4 blocks consisted of 2 microblocks (each contained 8 trials) and one calibration block. Participants wore glasses during blocks 2 and 4; they were without glasses in the remaining blocks. Letters denote trials presented repeatedly between consecutive microblocks.

were 4 types of repeated scan patterns: glasses-glasses (in Figure 1, they are denoted as C and G), noglasses-noglasses (A and E), glasses-noglasses (D and H), and noglasses-glasses (B and F).

Similarity of the repeated scan patterns was computed using Correlation distance

(CD). See Section 2.3.2 for details.

3.2.3 Data analysis

Analysis was done in a statistical program R (R Core Team, 2016). Within-subject ANOVA was used to test whether the similarity of scan patterns in the glasses-glasses condition differs from scan patterns in the noglasses-noglasses condition. Additionally, we tested whether the effect of putting glasses on or taking them off changed the scan pattern similarity (glasses-glasses vs. (glasses-noglasses / noglasses-glasses)). For computation of effect size, η_p^2 was used for within-subject ANOVA and Cohen's d was used for post-hoc tests. According to Cohen (1988), effect sizes were classified into small (Cohen's $d \sim 0.2$, $\eta_p^2 \sim .01$), medium (Cohen's $d \sim 0.5$, $\eta_p^2 \sim .06$) and large (Cohen's $d \sim 0.8$, $\eta_p^2 \sim .14$).

Blink detection

Because there are usually blinks in the recorded data, we needed to detect and exclude them. Blinks manifest as fast vertical movement with a decreasing pupil size. The eye tracker is sometimes able to detect them, but we employed our own procedure. Because velocity and acceleration can be mistaken for saccades, we decided to detect blinks using pupil size. For each trial, we computed a maximum pupil size and discarded all data with a pupil size less than 75% of this maximum. We also removed 30 ms before and after blinks to capture the start and finish of the blinks. Some blinks could have been missed using this method, but correlation distance is robust against the outliers. A total of 1,373 trials (79.46%) was included in the analysis; a remaining 315 trials were not included because of large amounts of blinks or technical errors.

Data preparation

For computation purposes, data was binned into the spatio-temporal grid where each bin had parameters $0.25° \times 0.25° \times 20$ ms.

3.2.4 Results

Tracking accuracy

Overall tracking accuracy was high (mean = 95%, SD = 12%). Accuracy did not differ between conditions ($F(3, 105) = 1.38$, $p = .254$, $\eta_p^2 = 0.02$). Because imperfect accuracy could result in different tracking strategies, we used only trials where participants correctly selected all four targets for further analysis. We also computed CD only for repeat trials; so we had 527 pairs of trials in total.

Calibration results

We tested whether calibration accuracy differed when participants wore glasses or not. Within subject ANOVA revealed no differences in average calibration error ($F(1,35) = 3.03$, $p = .091$, $\eta_p^2 = 0.03$) nor in the maximum error ($F(1,35) = 2.12$, $p = .154$, $\eta_p^2 = 0.02$). Calibration errors were lower for calibrations with glasses, as can be seen in Table 3.1.

	glasses	no glasses
Average error	0.42 (0.10)	0.45 (0.10)
Maximum error	0.88 (0.22)	0.93 (0.17)

Table 3.1: Means and SDs for calibration with and without glasses (in DVA). Standard deviations are shown in brackets.

Effect of glasses

There were significant differences between conditions ($F(3, 105) = 3.42$; $p = .020$; $\eta_p^2 = 0.05$). Tukey's post hoc tests revealed that differences between the glasses-glasses condition and the noglasses-noglasses condition were not significant ($p = .104$), but the difference between those two conditions is medium (Cohen's $d = 0.20$, 95% CI [-0.04, 0.45]). Tukey's post hoc test also revealed significant differences between noglasses-glasses and noglasses-noglasses conditions ($p = .002$; $d = 0.39$, 95% CI [0.15, 0.64]). As can be seen in Figure 3.2, mean CD in the noglasses-noglasses condition was lower than in glasses-glasses condition. When we added a number of diopters as a covariate, two-way ANOVA found the effect

of diopters to be non-significant ($F(1, 34) = 0.02$; $p = .890$) as well as the interaction with the condition ($F(3, 102) = 0.96$; $p = .412$).

Averaged CD for each participant did not correlate to the number of diopters ($r(34) = -.04$) , nor to the score from the FAcT-noglasses ($r(16) = .21$). However, this could be a chance finding due to the small sample size of data with the FAcT-noglasses value. Scatter plots for both diopters and FAcT can be seen in Figure 3.3. The number of diopters and results from FAcT were highly correlated ($r = -.70$; $p = .001$).

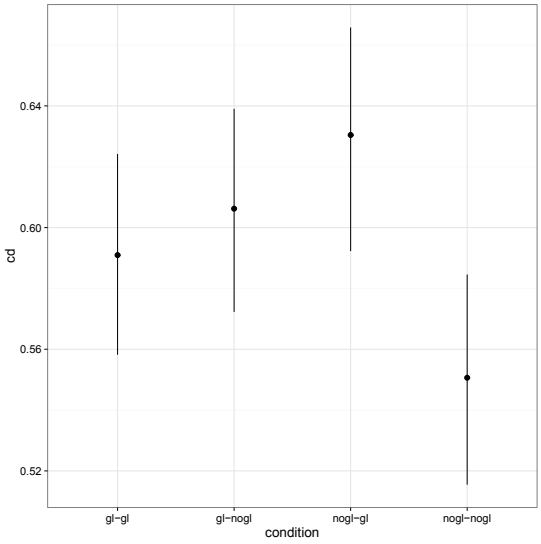


Figure 3.2: Correlation distances for each condition. Line denotes standard error of the mean. Higher values correspond to the higher overall distance (or lower coherence).

Interpretation of results in terms of scan pattern transformations

Average correlation distance for each condition could be interpreted in the terms of the results of scan pattern variability. The average correlation distance for all four conditions ranges from .56 (glasses-glasses condition) to .64 (noglasses-glasses condition). This would correspond to the spatial offset of 2–3 DVA, overall rotation of $> 25°$ or a scaling factor around .6 (see Tables A.1, A.2, and A.3 for reference).

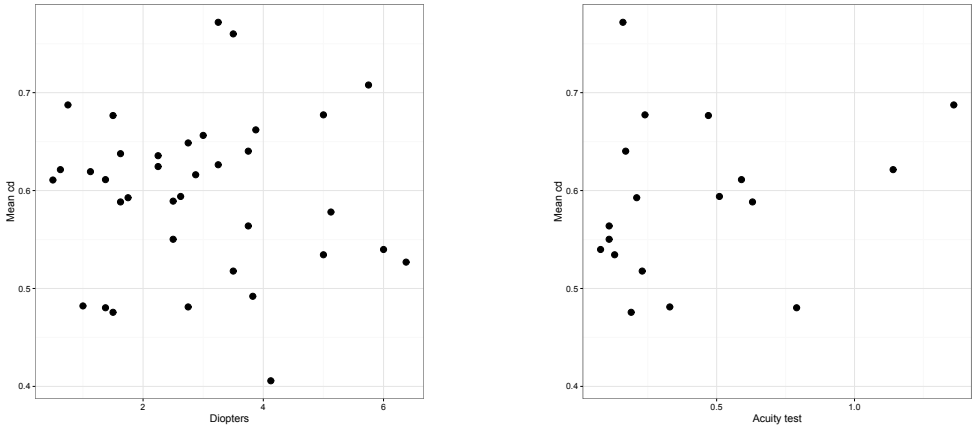


Figure 3.3: Scatter plot of mean CD per participant and number of diopters (left part) and FAcT-noglasses (visual acuity test measured without glasses).

3.2.5 Discussion

Differences between group of scan patterns

In this experiment, we showed one example of how to test whether two groups of scan patterns differ in overall coherence. In particular, we were interested in whether the scan patterns would be more coherent when we measured subjects with myopia: both with and without their glasses. This represents a common malpractice at some laboratories, when the administrator measures subjects without glasses in instances where he/she is having some trouble with the calibration procedure while participants have their glasses on.

A similar approach could be used for other dynamic tasks; in particular, the viewing of movie clips. Other metrics could be used for assessing the coherence of groups of scan patterns. For example, we could use average median distance between all scan patterns and compare this median distance between groups.

Our analysis could only answer, however, the question whether the conditions differ in overall coherence. We could not use this analysis if we are interested in how similar scan patterns are in one condition to those in other conditions. For example, when we removed the glasses, how did the scan patterns differ from scan patterns in other conditions. Even when we got identical coherence for each group, we did not know whether the scan patterns would be comparable. We offer a solution to this problem later on in this chapter.

Effect of wearing glasses on eye tracking measurement

Our results showed that, although we had a small sample, we were able to find differences in scan pattern coherence when participants wore glasses and when they did not. Therefore, researchers should be more careful when dealing with participants with myopia. In cases where administrators have trouble with calibration of subjects with myopia who wear glasses, they should have some spare eyeglasses on hand with good properties for eye tracking; i.e. low reflective properties of the glass in the lenses. In cases where the eye tracker cannot be calibrated with the glasses on, data from the measurements should be treated with care (if exclusion is not an option, analysis should at least be done with and without such problematic trials).

In addition, our results showed that wearing glasses resulted in a decrease of coherence for group scan patterns. This could be explained by the fact that, without glasses, they had to direct their eye gaze to the most informative place. While with the glasses, they could use different tracking strategies. The biggest difference was between conditions where scan patterns were presented twice without glasses and conditions where scan patterns were presented first with glasses and then without glasses. This difference could be the result of calibration, because there was no calibration between repeat presentations of scan patterns in the glasses-glasses condition. However, our main result concerning the difference between glasses-glasses and noglasses-noglasses conditions is not affected by this effect of extra calibration.

Errors for calibrating without glasses were slightly higher than with glasses, but the difference was not significant. This result is in compliance with manufacturer parameters for this particular eye tracker. If there were differences in the calibration procedure, we would not be able to compare eye tracking data from subjects with and without glasses.

Limitations

In this experiment, we tested the difference in coherence of scan patterns in MOT tasks. It would be beneficial to use a similar design for some static task; for example, for a visual search. In such static tasks, we could compare fixation

duration or saccade length between conditions with or without glasses.
To extend the results, it would be interesting to use eyeglasses with neutral lenses on subjects without myopia to see whether we get a similar pattern for subjects with normal vision. If there were in fact also differences between conditions, this difference would not be due to the effect of loss of visual acuity. Rather, it would be due to the effect of glass used for the lenses.

3.3 Methods for significance testing of group comparisons

In the previous experiment, we tested the overall differences between groups. The conceptually different question would be to test whether scan patterns in one group are more similar to each other than to scan patterns in the other group. In this section, we will denote two groups of scan patterns G_1 and G_2, each consisting of n_1 and n_2 scan patterns. scan patterns. The goal of this chapter is to develop new methods for testing whether scan patterns in each group are more different than scan patterns in other groups, and if they are more different than they are from each other. We will show an application of the methods in real experiments.

3.3.1 Feusner and Lukoff's approach for significance testing

To our knowledge, there is only one other approach for how to test differences between group of scan patterns. It is from Feusner and Lukoff (2008). Their approach is based on a permutation test and works as follows. First, we compute overall distance between groups (denoted as $d*$). Then we repeatedly divide scan patterns randomly into two groups and compute overall distance for that grouping (for repetition i we get distance d_i). Then we compare the $d*$ value with the 95% percentile of the distribution of d_i values, and we say that the groups are different if $d*$ exceeds the percentile. Overall distance is computed using the formula $d* = d_{between} - d_{within}$, where $d_{between}$ is the mean distance for

all pairs of scan patterns, where one scan pattern comes from the first group and the other from the second one (total of $n_1 \cdot n_2$ comparisons). Similarly, d_{within} is the mean distance for all pairs of scan pattern: both coming from the same group (total of $\binom{n_1}{2} + \binom{n_2}{2}$ comparisons). An arbitrary metric described in the previous chapter could be used for computing distance between two scan patterns. Tang, Topczewski, Topczewski, and Pienta (2012) extended this method for scan patterns with unequal length.

This algorithm can be speeded up by pre-computing distances for all pairs of scan patterns, and in each iteration just selecting the average distance for the given grouping. The metric is schematically captured in Figure 3.4. This method only works with distances between individual scan patterns and could not be employed when we use a metric computing the coherence of groups of scan patterns. We will call this method *pairwise comparison*.

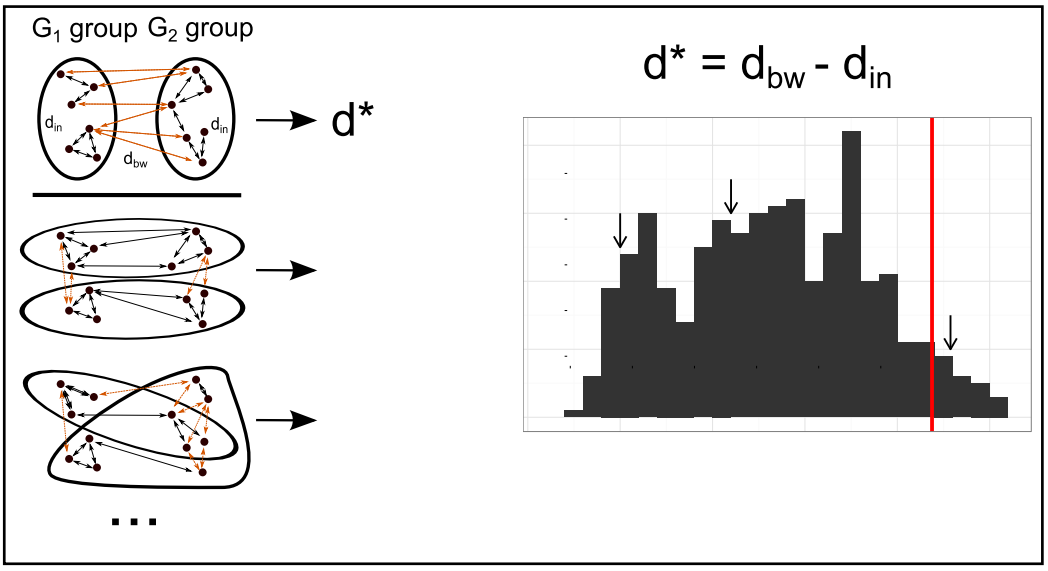


Figure 3.4: Method from Feusner and Lukoff for testing differences between group of scan patterns

3.3.2 Groupwise comparison

We created this method as an extension of the original pairwise comparison (method from Feusner and Lukoff). This method computes distance between whole groups instead of averaging the distance between all pairs of scan patterns. We call it a *groupwise comparison*, and it works as follows. We compute $d*$ for the

groups G_1 and G_2. Then we randomly divide scan patterns into different groups and compute the distance for that grouping. For robust estimates, the following inequality should hold $n_1 + \binom{n_2}{(n_1+n_2)/2} > 1000$, otherwise we will be selecting a quantile from a small sample. This schema is depicted in Figure 3.5.

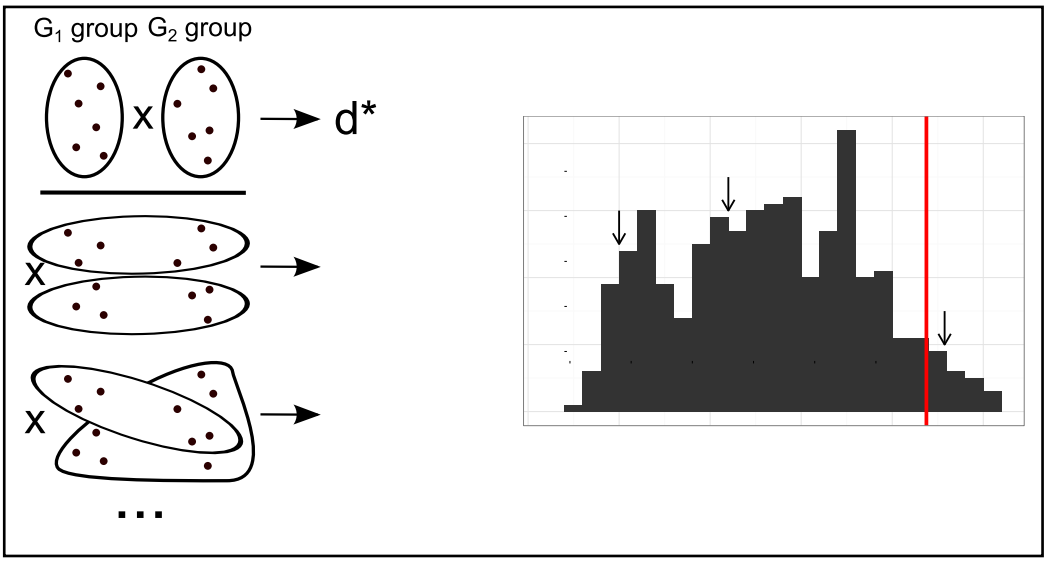


Figure 3.5: Groupwise method for testing differences between groups of scan patterns

3.3.3 Subset comparison

Subset comparison compares within-group variability for random subsets of merged G_1 and G_2 groups. If the scan patterns from each group are similar, we should get a similar within-group distance for the subset of scan patterns: irrespective of whether they were all selected from one group or whether they were selected from both groups. Therefore, using this strategy, we randomly sampled a subset of scan patterns ($n_s < min(n_1, n_2)$) and measured their overall distance. Then we compared whether this distance differed when scan patterns were selected from a single group or when scan patterns were selected from both groups. We denote the subsets of scan patterns from the same group as $\{G_i\}_{i=1}^{n_G}$ subsets and subsets of scan patterns from both groups as $\{M_i\}_{i=1}^{n_M}$ subsets. For G_i subsets, there are $\binom{n_1}{n_s} + \binom{n_2}{n_s}$ possibilities how to create subsets; for M_i, there are $\binom{n_1}{n_s/2} \cdot \binom{n_2}{n_s/2}$. The number of scan patterns forming each subset should be pre-selected to allow for multiple possible samples within each group. The number of scan patterns

forming G_i and M_i subsets should be selected appropriately, so both satisfy the following inequalities $n_G > 25$ and $n_M > 25$ for the results to be robust enough. The entire method is depicted in Figure 3.6.

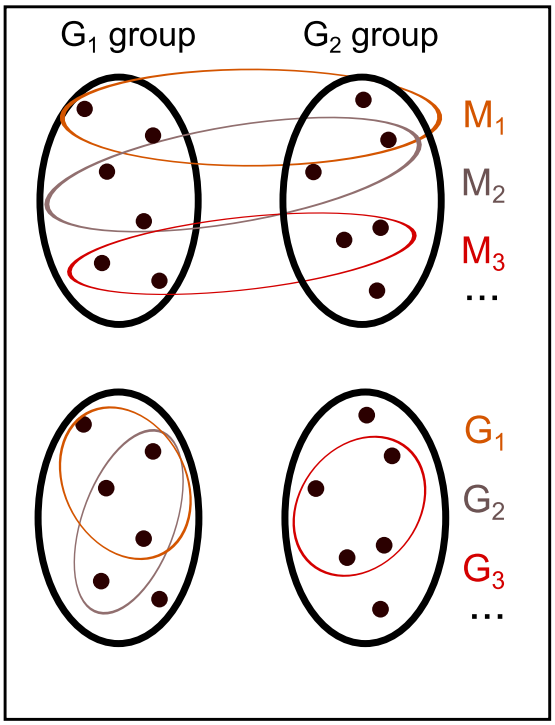


Figure 3.6: Subset comparison method for testing differences between groups of scan patterns

3.4 Simulation experiment 2 – Comparison of methods

We introduced two novel methods for testing differences between groups of scan patterns. It is unclear how sensitive the new methods are in comparison to the original pairwise method. To test their performance, we created artificial scan patterns, divided them into two groups and verified the capabilities of the three methods to detect correctly the ground truth division in the groups.

3.4.1 Method

We worked with artificial scan patterns in the experiment. Here, we will call them artificial trajectories, because they were generated as random walks, and

therefore they mimicked smooth pursuit movement and did not include simulated saccades.

Artificial trajectories

To verify the comparison methods for scan patterns with different variability, we decided to use artificial scan patterns that match the parts of scan patterns without saccades from an MOT experiment. To identify such parts, saccades were identified using an algorithm from Nyström and Holmqvist (2010). Intersaccadic intervals contained both smooth pursuit and fixations with varying numbers of raw samples. For each intersaccadic interval, we computed the average sample-to-sample distance and total length of scan pattern for each interval. Since those values were dependent on the number of samples in the intersaccadic interval, we used parts with 500–600 samples (corresponding to 2.0–2.4 s recorded with an Eyelink II at 250 Hz). We had 457 intervals satisfying this constraint.

Artificial trajectories were generated as random walks, in which each subsequent sample was created by sampling from bivariate normal distribution $(x_{i+1}, y_{i+1}) \sim N((x_i, y_i), \alpha \cdot I)$, where x_i and y_i is the last position, I is an identity matrix (2×2) and parameter controlling variance α varied from 0.0005 to 0.005 (with a step size of 0.0005). For the purpose of obtaining artificial trajectories similar to scan patterns without saccades, we set both x_1 and y_1 to zero. In addition to the varying parameter α, we also varied the smoothness of the trajectory by adding interpolated samples for each two subsequent generated points. The sparseness of interpolation was captured by the interpolation factor (F_i). This parameter varied from 1 (added zero interpolated samples between two generated samples using random walk) to 5 (added four interpolated samples) with a step size of .5. For each artificial trajectory, we also computed average sample-to-sample distance and total length. For the experiment, we selected parameters α and F_i which would generate artificial trajectories with similar properties as parts of real scan patterns (in terms of sample-to-sample distance and total length). For more robust results, we also selected parameters resulting in twice as variable scan patterns. The artificial trajectories consisted of 500 samples and to get a corresponding time scale the same as for the real data, we set the sample-to-

sample time difference to 4 ms (this was only important for the binning we will mention in the next section).

Final parameters with corresponding properties to artificial scan patterns are summarized in Table 3.2. For the experiment, we selected parameters $\alpha = .001$ and $F_i = 1.5$ (scan patterns with low variability) and $\alpha = .003$ and $F_i = 1$ (scan patterns with high variability).

	Artificial Data $\alpha = .003$, $F_i = 1$	Artificial Data $\alpha = .001$, $F_i = 1.5$	Real Eye Data
Sample-to-sample distance	0.07 (0.04)	0.03 (0.03)	0.03 (0.02)
Total length	34.27 (0.81)	13.18 (0.37)	15.11 (9.72)

Table 3.2: Properties of artificial scan patterns. Average sample-to-sample distances and total length of trajectories both with similar variability to real scan patterns (without saccades) and high variability. Values are shown for artificial trajectories consisting of 500 samples and 500–600 for parts of the human scan pattern.

Design

We used the following setting for evaluating the comparison methods. We repeatedly generated two groups of artificial trajectories; each contained 6 trajectories. Initial points for generating trajectories were positioned on the separate circles with a 1° diameter (each circle for each group)[2]. The distance between circles was constant (10°). The initial points in the G_1 group were positioned on the odd multiples of $\pi/6$ in the first circle, and on the even multiples of $\pi/6$ in the second circle for the G_2 group (as depicted in Figure 3.7). For each trajectory, we created additional identical copies and moved their spatial coordinates in the direction of the arrows from 0° to 5° (step size of 0.5°), so the radii varied from 1° to 6°. This manipulation gradually changed the initial obvious grouping into a less apparent one. For the spatial offset $> 5°$, the circles started to overlap. We generated the trajectories randomly alongside their copies 50 times; for scenarios with both more and less variable artificial scan patterns.

[2]We remind readers that the degree sign indicates the degree of visual angle.

We evaluated the three above-mentioned comparison methods in this setting. We used correlation distance for measuring distance between both individual scan patterns (for pairwise comparison) and groups of scan patterns (for groupwise and subset comparison). The data were binned in the 3D spatio-temporal matrix with a bin size of $0.25° \times 0.25° \times 20$ ms, so the number in each bin represented how many samples of the trajectory fell into the bin with the index corresponding to the sample coordinates. Regarding the correlation distance, we used a Gaussian filter with the parameters $\sigma_x = 1.2°$, $\sigma_y = 1.2°$ and $\sigma_t = 26.25$ ms as done in previous work (Dorr et al., 2010). Those values allows two scan patterns with high spatial variation to be treated as highly similar (almost 5 successive bins still have less than 1 standard deviation); while in the time scale, scan patterns should be aligned more tightly.

For the subset comparison, we used subsets of the groups containing 4 out of 6 scan patterns. For both pairwise and groupwise comparison, we would have had more than 900 possible groupings for the permutation test. For the given radius of the groups, accuracy was measured as a percentage of all correct rejections of null hypotheses that groups are random. Since the distance between groups and the radii of the groups were set to arbitrary values, we report the ratio of group over distance between groups in which strategies reached the chance level when rejecting the null hypothesis.

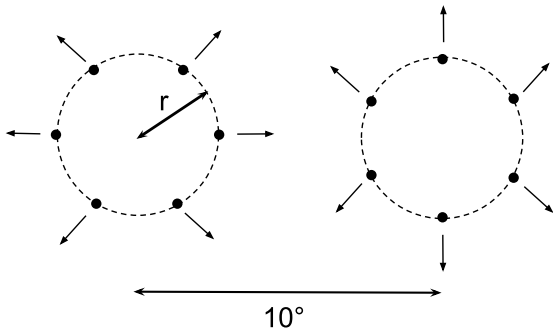


Figure 3.7: Schematic design for the comparison of the methods. Black dots represent initial starting points for each trajectory. Additional copies had their starting points moved spatially in the direction of the arrows.

3.4.2 Results

For all three comparison methods, accuracy decreased with the increased radius of the groups. Similar behavior was observed for both trajectories with high and low variability. The best discrimination capabilities were exhibited by subset comparison, followed by groupwise and pairwise comparison. When the circles were separated by more than 3°, all comparison methods were able to identify the grouping correctly. For distances between circles lower than 1°, all three methods were unable to identify the grouping. In other words, if we report the group variability as a percentage of the initial distance in which the methods started to discriminate differently, all three methods discriminated the groups at the distance of 143% of group variability or more; they failed to discriminate the grouping for distances lower than 111%.

We also fitted the data with a cumulative Gaussian filter to obtain threshold values. For less variable trajectories ($\alpha = .001$), the subset method reached chance level when the ratio of group radius over distance from center reached .41. The groupwise method reached chance level for ratio .40, and the pairwise method for the ratio .39. For more variable trajectories, results were similar: .43 for subset method, .41 for groupwise method, and .38 for pairwise method. As visualized in Figure 3.8, the decrease in accuracy is steeper for trajectories with lower variability.

3.4.3 Discussion

Our simulation experiment showed similar discrimination capabilities for all three strategies. The original pairwise comparison had the lowest precision. Out of the three methods, the pairwise method is the only one that compares each scan pattern against one another. The remaining two compute distances between whole groups of scan patterns. The same ordering of performance for the comparison methods was found for trajectories with both lower and higher variability. Although we used artificial scan patterns, our findings tend to generalize beyond that. Our artificial scan patterns were selected to show high resemblance in behavioral data, and we obtained similar results for scan patterns with low and high variability. Our method is also a general comparison strategy. Therefore, it

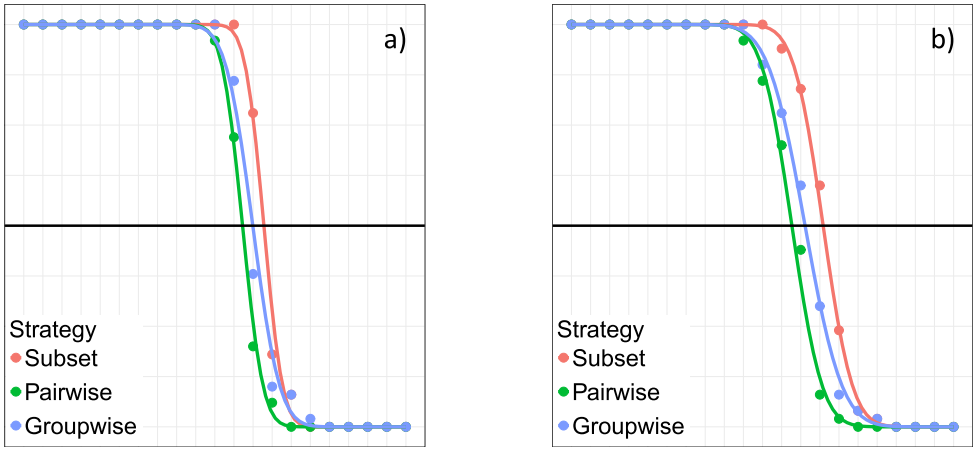


Figure 3.8: Accuracy of each of the three methods. Data were fitted by cumulative Gaussian filter. The decrease in the accuracy is more steep for trajectories with low variability (a) than for those with high variability (b).

works irrespective of the distance metric used for scan pattern comparison. There are some limitations to our study. First, we checked only a scenario where the two groups differ in their spatial location. In general, two groups of scan patterns could differ in other parameters, such as general shape. We believe that for basic assessment of the comparison methods, this approach is sufficient. Second, our two methods are not faster as concerns computing, which was one of the problems with the original metric from Feusner and Lukoff (2008). Although speeding up computation would be beneficial when using the methods for comparing groups of scan patterns in actual experimental designs, access to fast computers is not an issue anymore.

Our new methods can be used for general time series. The only requirement is to have a metric that computes group coherence for time series.

3.5 Experiment 3 – Left-right symmetry

In the previous sections, we described possible approaches to group comparison. Here, we show its practical application in the Multiple Object Tracking experiment. As we stated in the previous sections, MOT is a good alternative for studying scan patterns in dynamic tasks. This is because it requires sustained

attention for the tracking which leads to more similar scan patterns (both intra- and inter-subject). You can also present trials repeatedly without noticing.
We targeted for two goals in this experiment. First, we wanted to test whether we can geometrically transform trials to hide the repetition of identical trials. Second, we wanted to test the symmetry of scan patterns across left-right and top-down visual fields.

3.5.1 Introduction

As in our previous experiments, we presented some MOT trials repeatedly. In this case, we flipped half of the trials around the x- or y- axis. We could use this type of operation only case where scan patterns were symmetrical around the axes. This would mean that visual perception (and tracking in particular) did not differ in the left and right visual fields (or the upper and the lower visual field in the case of horizontal transformation).
Current literature is unclear about the existence of asymmetry. Left-right asymmetry is rarely reported among healthy subjects (Greene, Brown, & Dauphin, 2014; Petrov & Meleshkevich, 2011; Corballis, 2002), but there is evidence for upper-lower asymmetry (Levine & McAnany, 2005). In MOT, there have been several approaches to how eye gaze can be predicted based on scene content. If eye gaze is predicted only based on object position; such as tracking the center of the virtual polygon consisting of the targets (Yantis, 1992), there should be no differences between scan patterns. Since scene, in the MOT, is without any predefined orientation, there should not be any biases in the saccades‘ direction of amplitude during tracking (Foulsham & Kingstone, 2010). For the beginning of tracking, there are findings on preferences for early fixations to the left part of the scene (Dickinson & Intraub, 2009; Foulsham, Gray, Nasiopoulos, & Kingstone, 2013; Nuthmann & Matthias, 2014; Ossandon, Onat, & Konig, 2014). This fixation bias is usually discussed in relation to ”pseudoneglect”: a leftward bias in a line-bisection task in healthy humans (Bowers & Heilman, 1980; Jewell & McCourt, 2000).
In the task of free viewing of natural scenes, people tend to make more horizontal than vertical saccades, and there is no left-right asymmetry in saccade orientation

(Foulsham, Kingstone, & Underwood, 2008). For an antisaccade task, people are usually more prepared to make rightward saccades than leftward ones. They also exhibit fewer errors and make those rightward saccades faster (Evdokimidis et al., 2002; Tatler & Hutton, 2007). However, this asymmetry may not be evolutionarily coded, but rather it might be the result of a learned behavior. Abed (1991) showed that directions of saccades (when looking at a simple display with dot pattern) are also biased differently for Western, Middle Eastern and East Asian participants.

Static stimuli were used in all of the above-mentioned studies. Presented asymmetries in fixation location happened only at the beginning of the trial, and they disappeared later (Nuthmann & Matthias, 2014; Ossandon et al., 2014). To our knowledge, there are no other studies regarding the symmetry of eye movements using dynamic stimuli. It is an open question as to whether there will be left-right asymmetry in the scan patterns in a dynamic task like MOT. Before each tracking portion of the MOT, there is a static part in which participants memorize the identity of targets. There is the possibility that leftward biases will disappear during tracking.

We will first describe the experiment regarding left-right symmetry: denoted here as Experiment 3.

3.5.2 Method

Participants

Thirty-one students (27 females; ages 19–28, mean 20.8) participated in the experiment in exchange for course credit. All of them had normal or corrected-to-normal vision (wore glasses or contact lenses). None of them had participated in this type of experiment before. Originally, we collected data from 32 participants, but we had to exclude one due to a technical error.

Apparatus and stimuli

The apparatus and stimuli were identical to those in Experiment 1.

Procedure

The procedure was identical to Experiment 1. Each participant completed 90 trials divided into six blocks (each block consisted of 15 trials). There were five extra training trials before the experiment. Those were not, however, included in the analysis. In each block, the fifteen trials were divided into the following segments: five experimental trials (*L trials*); and five trials, in which the object trajectories from the L trials were flipped around the y-axis (*R trials*). The remaining five trials were unique per each block, and they were added to reduce chances participants would notice repetition. Figure 3.9 depicts one frame of the L and the corresponding R trial with direction of movement for each object.Therefore, we presented the same trial six times under the normal condition and six times under the flipped condition. Again, we generated a 10 s period of motion, which randomly started at some time point between 0 s and 2 s. In this case, each trial lasted 8 s, therefore we had 6 s of the common movement that all trials shared in the worst cases of overlap. The entire experiment lasted approximately 45 min.

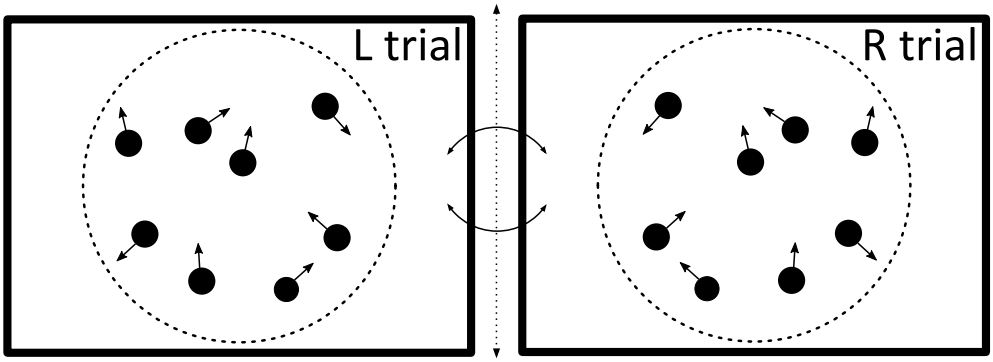


Figure 3.9: Illustration of the flipping of object trajectories for one particular frame for normal trials (L trial) and their corresponding flipped variant (R trial). Each object has its direction visualized using a small arrow.

3.5.3 Data analysis

Blink detection

Blink detection was done similar to Experiment 1.

Data preparation

We discarded all of the eye gaze data outside the range $-15°$ to $+15°$ in both vertical and horizontal directions. Objects moved in a circular arena, and therefore some of the eye gaze data outside the arena could be missed by this method. Such data constituted only 0.07% of the samples, and so we decided to retain that data for analysis. Similar to previous experiments, we binned data into a 3D spatio-temporal matrix with bin size $0.25° \times 0.25° \times 20$ ms.The scan patterns from R trials were flipped around the y-axis to ensure scan patterns comparable to L trials.

Distance metric and comparison method

Again, we used correlation distance for computing the similarity of groups of scan patterns. Because in our simulation experiment subset comparison scored the best out of the three, we used only this method for testing significance. For each repeated presentation of the same trial, we had two groups of L and R trials which correspond to the groups G_1 and G_2 from the simulation experiment. Each group represented scan patterns Each group represented scan patterns from the repeated presentations in the same trial. Therefore, we had five pairs of groups of scan patterns for each participant. Again, subsets were formed from four out of six scan patterns (each L and R group had 6 repeated presentations of the same trial).

3.5.4 Results

Overall, tracking accuracy was high. All four targets were correctly selected in 91% of all trials. Per-subject accuracy ranged from 76% to 99%. Because tracking strategy is dependent on which objects participants consider to be targets, trials where some objects were incorrectly selected as targets resulted in different scan patterns. Therefore, we selected only trials where all four targets were correctly selected.

Differences between L and R trials were tested using linear mixed models with Subject ID and Trajectory ID as random factors (the same trajectory was presented in L and R trials). It is still an open question as to whether p-values

should be computed in the case of linear mixed models. So, we decided to test the significance with classical model comparison using an χ^2-test. Similarly, there is no consensus about effect sizes for linear mixed models. So, because we are interested in differences between two groups, we will show the relative difference as a percentage relative to the baseline.

Using subset comparison, we had three similarity conditions: L (similarity within subsets of L trials, the original trajectory), R (similarity within subsets of R trials, the flipped trajectory), and M (similarity within subsets of both L and R trials). If the transformation did not induce any additional variance in scan patterns, both the L and R groups should be treated as just random groupings of scan patterns without any statistical difference. There could also be some systematic difference introduced by the flipping. However, because the division into the groups was arbitrary (we could reverse the direction and see the L trials flipped to R trials), this difference would not project into the differences. Thus, crucial comparison would be between the averaged values for L and R versus the M value. The potential difference showed that scan patterns in flipped trials differed significantly from the original trial. We denoted mean value from the L and R values as the LR value (in the description of the subset comparison, this value was denoted by the subsets as G_i).

There was no difference between the coherence of L trials (mean = .47, SD = .12) and R trials (mean = .47, SD = .12) when using mixed models ($\chi^2(1) = 0.01, p =$ .920). The CD values for the subsets formed by both the L and R trials were significantly higher[3] (mean = .53, SD = .11) than the average from the LR values ($\chi^2(1) = 89.4, p < .001$). In terms of effect size, the M value was only 13% higher than the LR value. We aggregated CD values per each trajectory (we therefore got one LR and M value per each trajectory) and correlated the aggregated values. The M and LR strongly correlated ($r = .83, p < .001$) showing that when we mix scan patterns from the flipped and normal condition and compute coherence for subgroups, it has similar variance to scan patterns from either condition.

[3]Note that we are using distance. Therefore, larger values denote more distant groups.

How did the scan patterns differ?

We found significant differences between the original and flipped trials. It is unclear whether it is due to difference in shape, some spatial offset or some more complex difference. In our case, there was a potential source of error in the non-centered viewpoint (imprecise placement of the chinrest). To verify, that this was not the case, we systematically varied the overall position of scan patterns from R trials and computed their similarity to the L trials. For each scan pattern from R trials, we moved the x coordinate from $-0.5°$ to $+0.5°$ with a step size of $0.25°$ relative to the original x position. Even after this simple manipulation, the differences remained significant ($p < .001$). Therefore, the difference between the L and R groups was not due to spatial shift. We also tested shifting the R trials for each block separately (same x-shift range as before but for each block uniquely, therefore 6^6 combinations), but the differences were still significant ($p < .001$). This manipulation tested whether spatial offset could be introduced through differences in the calibration procedure at the beginning of each trial. In terms of scan pattern variability, the average correlation distance for L and R trials corresponded to the spatial offset $< 2°$.

3.5.5 Discussion

We found out that when we flip trials around the y-axis, the scan patterns are different from those in the repeated presentations in the same trial. In terms of effect size, this difference was small (only a 13% increase in the overall distance), and therefore we could use this technique for the masking purposes of the repetition. More specifically, when presenting trials repeatedly, we could present half of the trials with flipped trajectories and then flip them back before the analysis. Eye movements in MOT are usually studied for the purpose of tracking strategy (see Chapter 4). Combining trials with their flipped copies would introduce extra noise, but this increase in noise is highly correlated with the coherence of groups from repeated trials. Only because of this, would we get qualitatively similar results when designing strategies. This means that we can use a smaller number of trials for masking purposes, and thus can we include more experimental conditions in one session (with respect to the time constraint of each session and

participant tiredness).

Our experiment showed an application for one of our methods for the group comparison. We were able to express the difference between the groups as an increase in overall distance (or decrease in coherence). For other research areas, a similar approach could be used to test differences between groups and decide the threshold for the decrease in coherence that is acceptable for the given purposes.

3.6 Experiment 4 – Upper-lower symmetry

Similar to the previous experiment, we wanted to test similar techniques for the masking and flipping of the trials around the x-axis. Again, we presented some trials repeatedly while presenting a flipped version in some cases. We wanted to test whether scan patterns from flipped trials were similar to scan patterns from the non-flipped condition.

3.6.1 Method

Participants

Thirty-two (27 females, ages 19–28 years, mean 21.8) students participated in this experiment in exchange for course credit. All of them had normal or corrected-to-normal vision. None of them had participated in this type of experiment before.

Apparatus and stimuli

The apparatus and stimuli were the same as in Experiment 3.

Procedure

The same procedure was used as in Experiment 3. The only difference was the experimental manipulation; here, we flipped trials around the x-axis. The unmodified trials are denoted as *U trials* (upward) and flipped ones as *D trials* (downward).

3.6.2 Data analysis

We again used the CD metric for evaluating distance between groups and a subset strategy for testing the significance of the comparison. We denote the average value of U and D trials as the *UD* value.

3.6.3 Results

Similarly to Experiment 3, the overall tracking accuracy was high. Participants correctly selected all four targets in 96% of all trials, and per-subject accuracy ranged from 86% to 100%. Only trials where all four targets were identified were included in the analysis.

Regarding differences between trials, we found no difference between U trials (mean = 0.49, SD = 0.14) and D trials (mean = .48, SD = .14) using linear mixed models ($\chi^2(1) = 0.68, p = .409$). The CD value for mixed UD trials was significantly higher (mean = .59, SD = .14; $\chi^2(1) = 139.93, p < .001$). Again, on the trajectory level, UD and M values strongly correlated ($r = .82, p < .001$). In terms of relative increase of distance, the increase in the M values was 22% relative to the UD values[4]. We did not perform additional manipulation of the D trials, i.e. such as varied spatial offset for the y coordinate.

3.6.4 Discussion

Similar to Experiment 3, we found significant differences between scan patterns from repeated trials and those with the flipped variant around the x-axis. The difference was larger than in Experiment 3 (22% for Experiment 4; 13% for Experiment 2), confirming that there is greater asymmetry between upper and lower visual fields than there is between left and right visual fields. We can expect the extent of dissimilarity between horizontal and vertical planes, because there are a lot studies supporting this claim such as those on visual acuity (Freeman, 1980), mental imagery (Finke & Kosslyn, 1980), and the extent of crowding (Toet & Levi, 1992).

[4]This value was 13% for the Experiment 3.

3.7 General discussion (Experiments 2 – 4)

In this chapter, we introduced two novel methods for statistical testing of the significance as to whether two groups of scan patterns differ. We evaluated the performance of the methods and compared their performance for discriminating between two groups of artificial scan patterns. Finally, we tested one of the methods (the one with the best performance on artificial data[5]). There are several implications from our results.

3.7.1 Group comparison of scan patterns

The majority of studies employing eye tracking use static stimuli. In many cases, this is appropriate for answering research questions, but sometimes the dynamic stimuli increase the amount of information we can mine from data. For example, Bonnen, Burge, Yates, Pillow, and Cormack (2015) developed a tracking task for estimating psychophysical curves that generates in a one-minute session the same amount of data as does the classical approach for signal detection theory. Because there is currently only one method for testing the significance of differences between groups of scan patterns, researchers are limited in their analyses to comparing within-group coherence only.

3.7.2 Methods for masking the repetition in MOT

Typically, Multiple Object Tracking is used to study divided attention. We encourage researchers to use MOT to study eye movements in dynamic tasks. Due to the nature of the task, it has some advantages over movie clips. We can easily present trials repeatedly to study the intra-subject coherence of eye movements. Because participants' performance gets worse with increased fatigue, the length of the experimental session is limited. Therefore, there are several options for how to reduce the chance of noticing repetition.

- Add filler trials to the experimental protocol. This is a simple solution, but it would lead to longer experiments.

[5]It is important to note that the performance of all three methods was similar; the subset comparison scored the best out of the three.

- Randomize the start of the trials. We could generate trials in advance and randomize the starting time for each of them.

- Present modified trials and apply the inverse transformation on the scan patterns. This technique that we introduced in this thesis slightly reduces the coherence of the scan patterns, but the relative difference is small (13% for the flipping around y-axis, 22% for the x-axis).). As the mean coherence for the group of scan patterns from both the original and modified versions is highly correlated ($r = .83$ for y-axis and $r = .82$ for the x-axis), we could use this technique for masking.

3.7.3 Asymmetry of scan patterns

The final findings from Experiments 3 and 4 relate to the perceived asymmetry of scan patterns. The differences between original and flipped trials could come from different sources. First, there are asymmetries in the smooth pursuit movements, which are the main part of scan patterns in MOT (Ke, Lam, Pai, & Spering, 2013). Second, human visual perception is not symmetrical across the visual field (Bradley, Abrams, & Geisler, 2014; Najemnik & Geisler, 2009). Therefore, humans try to compensate for the asymmetries by changing both tracking strategies and eye movements. Finally, there are some learned biases such as reading direction or central bias (Tatler & Hutton, 2007) that could affect the shape of the scan pattern.

This asymmetry is not reflected in the current eye movement models in the MOT (Fehd & Seiffert, 2008, 2010; Lukavský, 2013; Děchtěrenko & Lukavský, 2014). These asymmetries also contrast with the idea that people track targets as single objects and fixate on the centroid of the group (Fehd & Seiffert, 2008; Foulsham & Kingstone, 2010; Yantis, 1992).

Differences in gaze patterns were greater for the vertically flipped trajectories than for the horizontally flipped ones. This is in accordance with studies showing lower-upper asymmetry in the visual field (Greene et al., 2014; Hagenbeek & Van Strien, 2002; Pitzalis & Di Russo, 2001) and left-right asymmetry (Nuthmann & Matthias, 2014; Ossandon et al., 2014). The differences between repeated presentations in the same trial were complex. So, overall coherence could not be

improved by adding overall spatial offset to the scan patterns.

3.7.4 Limitations

There are several limitations for our study. First, we did not fully test discrimination capabilities for the group comparison methods in our Simulation experiment 2. We only tested the case where the correct answer was that two groups differ. We would have gotten better evaluation of the methods, had we created a paradigm in which metrics from signal detection theory could be employed; such as a discriminability index d' (Green & Swets, 1966). Second, we only tested the possibility of flipping the stimulus in one particular task. Further exploration should be done to test this idea in other dynamic tasks and in movies in particular. Finally, our two methods are still computationally demanding. Therefore, we still have a long way to go before the analysis of group differences will be of widespread use.

Although our approach is focused on the use of the MOT task, we could generalize our findings for tasks in which both top-down and bottom-down influences are controlled (as in the MOT task).

4. Machine learning in MOT task

In previous chapters, we introduced MOT as a task useful for studying the properties of scan patterns. The main advantage of MOT is the influence of both the bottom-up and top-down processes for planning eye movements. Therefore, when using MOT to study the parameters of scan patterns, it is beneficial to be able to explain the fixated position with respect to the objects in the scene. This would allow us to specify the scan patterns as time series with individual parameters corresponding to the objects in the scene.

In this chapter, we address traditional research questions regarding eye movements in MOT. In particular, we are interested in predicting eye gaze using a data-driven approach. We used feed-forward neural networks for this purpose.

4.1 Experiment 5 – Neural network modelling of eye movements

So far, several different models have been proposed for predicting eye gaze position in MOT (see Section 1.6.3). However, none of the models performed as well as scan patterns from the repeated presentation in the same trial. In other words, if we want to predict eye movements for a specific MOT trial, we would get better results by using the scan pattern from another presentation of the same trial. Although we could come up with other models, we would reach the limit introduced by the variability of the scan patterns. To test whether there is still a possibility for better strategies, we used eye gaze data from three different MOT experiments and trained feed-forward, multilayer perceptron using this dataset to predict scan patterns in MOT. Because scan patterns in MOT are variable, we used the coherence of the group of scan patterns from repeated presentations as the baseline. This experiment builds on work from Děchtěrenko (2012). It is an extension of the author's paper (Děchtěrenko & Lukavský, 2016).

4.1.1 Methods

Data description

We used data from three MOT experiments (two of them were described in Děchtěrenko and Lukavský (2014), the third in Chapter 3.5). To avoid confusion, we will denote the experiments from Děchtěrenko and Lukavský (2014) as Experiments 6 and 7. For brevity, we do not include the description of the experiments, because it is similar to the experiments presented in this thesis. We only describe the differences in design that we employed in this thesis.

The datasets consisted of eye tracking data for a total of 42 participants (9 males, mean age = 21.16). All of them had normal or corrected-to-normal vision. Experimental stimuli and the procedure were the same as in Experiment 3 described in this thesis. Each participant was presented 40–95 trials (95 trials were presented in experiments 2 and 4; 40 trials in Experiment 7). Objects moved at a speed of $2°/s$ – $5°/s$.Participants viewed some trials repeatedly in all three experiments. In Experiments 6 and 7, there were trials presented four times; in Experiment 3, the experimental trials were presented six times.

Since the experiments were designed to test different things, there were differences between the individual experiments. In Experiments 6 and 7, objects moved at a speed of $2°/s$ – $2.2°/s$. It differed for each participant based on their initial tracking abilities. In Experiment 2, they moved at a speed of $5°/s$. In Experiments 6 and 7, objects moved in a rectangular arena with sides 30° in length (for comparison, in Experiment 3, objects moved in a circular arena) and in those two experiments, objects did not bounce off each other as in Experiment 3. The only difference between Experiments 6 and 7 was in the number of presented trials, because Experiment 7 was an exact replication of Experiment 5. For a detailed description of Experiments 6 and 7, see Děchtěrenko and Lukavský (2014). We summarized all differences in Table 4.1. Since we wanted to generalize our predictions for a wider array of MOT experiments, minor differences between presented experiments were not crucial in our case.

Parameter	Dataset 1 (Experiment 3)	Dataset 1 (Experiment 6	Dataset 1 (Experiment 7)
No. of pariticpants	31	8	3
Trials per subject	95	95	40
No. of repeated trials	6	4	4
Speed	$5°/s$	$2°/s$–$2.2°/s$	$2°/s$
Arena	Circular 30° in diameter	Rectangular 30° one side	Rectangular 30° one side
Bounce style	each other arena borders	arena borders	arena borders

Table 4.1: Differences between experiments. All other parameters of the experiments were the same.

Data preprocessing

First, artifacts such as blinks were removed from the data (see Section 3.2.3). After the exclusion of artifacts, we had more than 1,000,000 potential samples which corresponds approximately to the 3.8 hrs. of tracking. We wanted to compare the predictions to the intra-subject variance, so we selected only trials presented repeatedly. In Experiment 6, the number of distractors differs based on the experimental condition. Therefore, we selected only the repeated trials with four targets and four distractors. The data from the repeated presentation of the identical trial would differ only in eye gaze position; the position of the targets and distractors would be identical. Thus, for each time stamp and repeated trial, we selected the eye gaze data most proximate to the eye gaze's median position (measured using Euclidean distance). To get insight into the variability of eye gaze, we presented the x coordinate from one trial in the dataset in Figure 4.1. Because the sampling rate of the eye tracker is 250 Hz, the time difference between consecutive samples would be only 4 ms and the dataset would be highly autocorrelated. Therefore, we took only each tenth sample from each trial. Since objects moved at speeds of 2°/s–5°/s, the distance of each object between two consecutive samples would be on average 0.2°–0.5°. Therefore our final dataset had 48,000 samples (1,534 trials or 2.5 hours of tracking time). Our dataset was

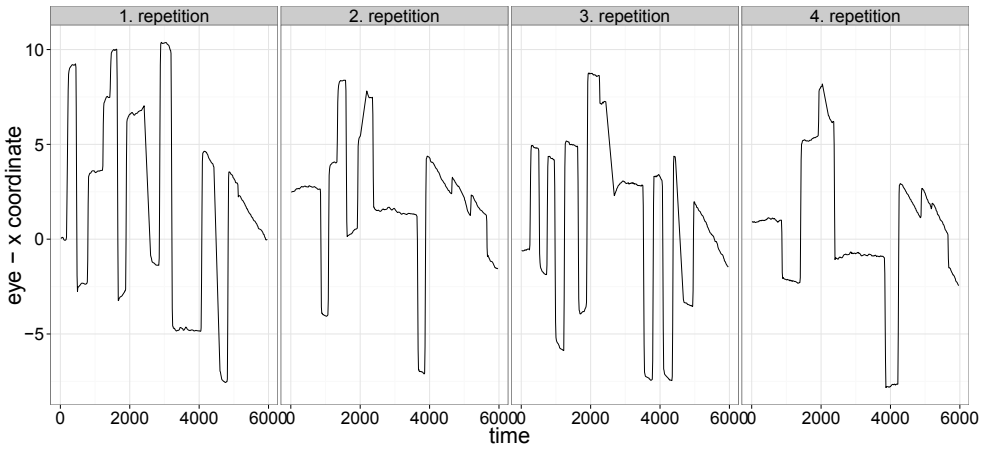


Figure 4.1: X coordinate over time for one repeated trial.

in the format $x_1, x_2, \ldots, x_8, y_1, \ldots, y_8, eye_x, eye_y$, where x_i,y_i were x and y coordinates for object i and eye_x and eye_y were eye gaze coordinates for a particular frame. First four objects ($i \in \{1, 2, 3, 4\}$) were targets and the rest were distractors. At the beginning of the trial, we could swap the order of the targets in the dataset and we would get same results. This is because the objects (within the set of targets and distractors) were indistinguishable. Therefore, we should get the same eye gaze position for the inputs $x_1, x_2, \ldots, y_1, y_2, \ldots$ and $x_2, x_1, \ldots, y_2, y_1, \ldots$ and similarly for other combinations of targets and distractors (but without permuting positions of targets and distractors). Without including such symmetry in the dataset, we would bias the neural networks for specific configuration. One solution for how to proceed would be to add new samples by permuting targets and distractors individually. This would increase the size of the dataset dramatically, since we would get $4!4! = 576$ producing identical output for each sample. Another solution is to sort inputs to get unambiguous representation for each combination of target positions. Specifically, we sorted the data set by the x coordinate (individually for targets and distractors) and also sorted y coordinates accordingly. For example, if we have targets with an x coordinate (rounded to degrees for better intuition about the sorting) 4, 1, 3, 2 and with a y coordinate 12, 10, 11, 9, after sorting, we get vector 1, 2, 3, 4 for the x coordinate and 10, 9, 11, 12 for the y coordinate. We also sorted each coordinate separately. In this case, object mapping was lost, but some of the strategies might be unaffected by this approach, if they predict eye gaze for each coordinate separately. Therefore,

we had three versions of the dataset:

- Unsorted – $x_1, x_2, \ldots, x_4, x_8, y_1, \ldots, y_8, eye_x, eye_y$
- Sorted by x coordinate – $\sigma^T(x_1), \ldots, \sigma^D(x_8), \sigma^T(y_1), \ldots, \sigma^D(x_8), eye_x, eye_y$, where σ^T (resp. σ^D) is the permutation ordering targets (resp. distractors) by the x coordinate
- Sorted for each coord. – $\sigma_x^T(x_1), \ldots, \sigma_x^D(x_8), \sigma_y^T(y_1), \ldots, \sigma_y^D(y_8), eye_x, eye_y$, where σ_x^T (resp. σ_x^D) is the permutation ordering targets (resp. distractors) by the x coordinate and σ_y^T (σ_y^D) by the y coordinate

Problem analysis

As our previous research showed (Děchtěrenko, 2012), the number of neurons in the hidden layer did not have an influence on the quality of prediction. We also experimented with different types of neural networks (recurrent neural networks) and machine learning algorithms in general (ϵ-SVM), but the simple feed-forward neural networks showed similar performance for this problem. The main limitation in the quality of prediction was the variability in the dataset, which we tried to decrease in the previous subsection; the type of machine learning algorithm did not cause limitations.

Neural network description

We used MATLAB with the Neural Network Toolbox (8.0.1) for the training and fitting of the neural networks. We used feed-forward networks with one hidden layer with 50 nodes, an output layer with two nodes (the predicted position of x and y coordinates of the eye gaze). We wanted to train networks using data with different types of information available, so we trained them using different feature vectors. The number of nodes in the input layer corresponded to the length of the feature vector. Random samples from the dataset presented to the input layer were non-linearly transformed via a sigmoidal transfer function to the nodes in the hidden layer and then via the same transfer function to the output layer. The training stopped, when we reached one of the criteria for stopping. We used default values for training. See toolbox manual for detailed

instructions. When propagating values from input to output, mean squared error was computed and weights for connections between layers were adjusted using the Levenberg-Marquardt back-propagation algorithm (Marquardt, 1963). All data were rescaled to the range (-1, 1) before training and then divided into the training, validation and test sets at a ratio of 70:15:15. The entire scheme is depicted in Figure 4.2. We used different parametrizations of the dataset;

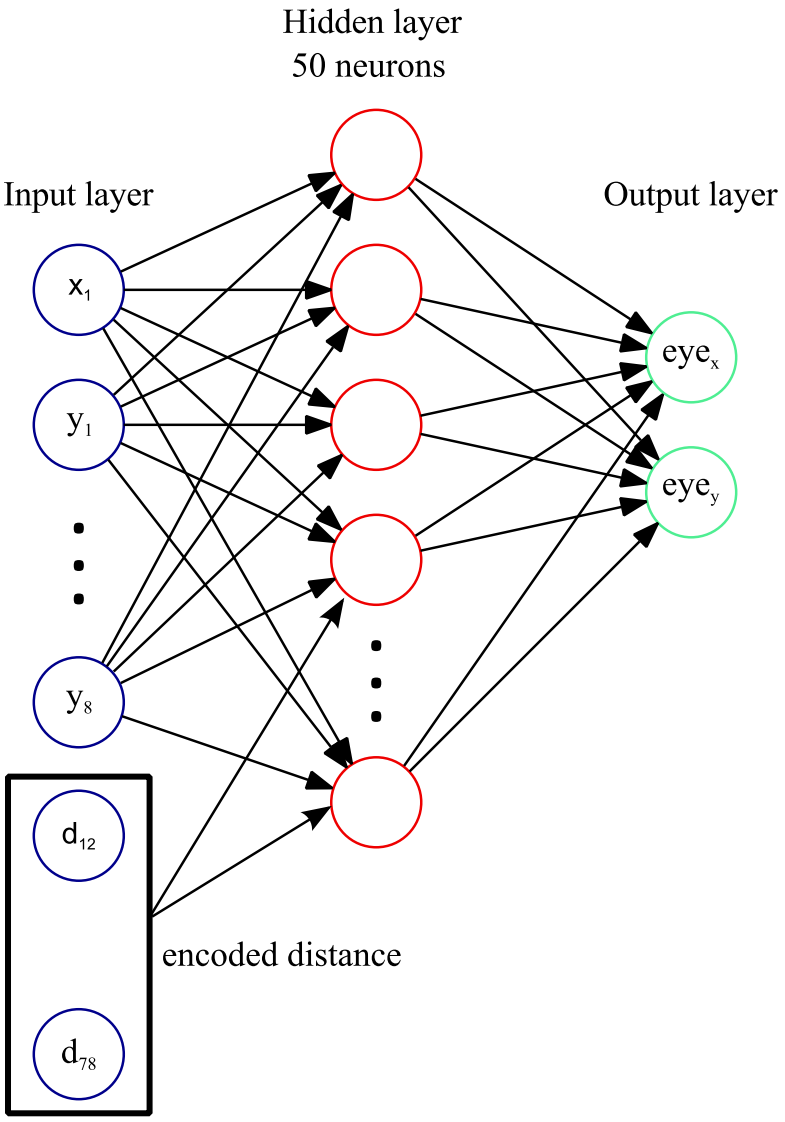


Figure 4.2: Scheme of the neural network used for predicting eye gaze position. Encoded distance between neurons is added into the model for several neural networks models. A sigmoid function was used as a transfer function.

each of them contained different information. In half of the models, we used the positions of all eight objects; in the other half, we used target positions only

(therefore, the feature vector had half the length). We also created two variants with distance matrix for each group. In one case, we added distance between targets and distractors (additional $4 * 3/2 = 6$ inputs) and between all objects (additional $8 * 7/2 = 28$ inputs). Finally, we also tested a model which used all eight objects, and we added the position of the eye gaze data from the previous frame. This model was qualitatively different, since it used information from other frames and not from the current frame only. All models are summarized in Table 4.2. This parametrization allowed us to compare models trained using

Objects used	Distance added	Feature vector
Targets only	No distance	$(x_1, \ldots, x_4, y_1, \ldots, y_4)$
	Target-Distractor	$(x_1, \ldots, x_4, y_1, \ldots, y_4, d_{1,5}, \ldots, d_{4,8})$
	All objects	$(x_1, \ldots, x_4, y_1, \ldots, y_4, d_{1,2}, \ldots, d_{7,8})$
All objects	No distance	$(x_1, \ldots, x_8, y_1, \ldots, y_8)$
	Target-Distractor	$(x_1, \ldots, x_8, y_1, \ldots, y_8, d_{1,5}, \ldots, d_{4,8})$
	All objects	$(x_1, \ldots, x_8, y_1, \ldots, y_8, d_{1,2}, \ldots, d_{7,8})$
Previous eye gaze	-	$(x_1, \ldots, x_8, y_1, \ldots, y_8, eye_x, eye_y)$

Table 4.2: Different feature vectors used pro prediction of the eye gaze.

data with different amounts of information. For example, a model with targets only had no information about the distractors. Therefore, if it scored similar to real eye gaze, we could assume that eye gaze is influenced by targets only.

Evaluation

We had three different variants of datasets (no sorting, sorted by the x coordinate and sorting of each coordinate individually) and 6 models for each of the datasets (and 1 model with information about previous eye gaze). This equals 19 models in total. We repeatedly trained each model on randomly-divided datasets to overcome local minima of the MSE function used for evaluation of the models. We computed Euclidean distance between predicted value and real eye gaze for each sample. The median value of the distances for the whole dataset was used for evaluating the model‘s performance.

As a baseline, we used the median distance between all pairs of eye gaze from

repeated trials. Therefore, if we had 4 repeated presentations, we computed the median from $\binom{4}{2} = 6$ values.

To anchor predicted values, we also predicted eye gaze using other analytical strategies and evaluated them in the same way as neural networks. We used the following strategies:

- *Center strategy* – eye gaze is predicted constantly to the center of the display (at coordinates (0,0)).
- *Target-centroid strategy* – eye gaze is predicted at the averaged position of the targets.
- *All-centroid strategy* – eye gaze is predicted at the averaged position of all objects.
- *Anti-crowding strategy*(Lukavský, 2013; Děchtěrenko & Lukavský, 2014) – eye gaze is predicted to the position minimizing the following ratio

$$\vec{x} = \underset{\vec{x}}{\operatorname{argmin}} \sum_{t \in T} \sum_{d \in D} \frac{\|\vec{x} - \vec{t}\|}{\|\vec{t} - \vec{d}\|}$$

  where T is the set of targets, D is the set of distractors and $\|\cdot\|$ is a norm of a vector.

4.1.2 Results

All results are summarized in Table 4.3. The median value of all pairwise distances was 2.59°. Of the analytical strategies, the Anti-crowding strategy had 22.78% larger values than the baseline, followed by Target-centroid, All-centroid and finally the Center strategy. All of the neural networks scored better than the All-centroid strategy. In general, sorting the dataset increased prediction. Surprisingly, prediction was increased even when we sorted each coordinate separately. This showed that increased variability of the sorted dataset was more important than correspondence of the coordinates to the object. Therefore, we could dissociate each coordinate when predicting eye gaze. This is plausible, because we could still compute mean strategies biased to the center. Since the computation of the distance between objects could be encoded in the neural network, adding information about the distances between objects did not improve

prediction. There is one important exception. For the model, where only target positions were used, but the distance matrix between all pairs of target-distractors was added, we achieved the best predictions. It was only 12.36% higher than baseline (almost 10% better than the best analytical strategy). This particular model is important in the sense that there was no information about the position of the distractors in the dataset. We only had information on the distance of the distractors to the target. As expected, predictions from the model using previous eye position scored better than the baseline (the median distance was only 0.27°). Predictions for one particular frame are visualized in Figure 4.3. In this particular case, neural network predictions were closest to the eye gaze in this trial. Overall, our results show that incorporating information about distractor location increased prediction of eye gaze.

Because datasets from each individual experiment were small, we did not test separately whether there were some differences in training using each dataset.

Baseline			Median distance (in °)	Rel. difference to baseline
Same subject, repeated trial			2.59	0%
Analytical strategies				
Center			5.55	+114.29%
Target-centroid			3.56	+37.45%
All-centroid			4.16	+60.62%
Anti-crowding			3.18	+22.78%
Neural networks				
Objects used	Sorting	Distance added		
Targets only	No sorting	No distance	3.54	+36.68%
		Target-Distractor	3.45	+33.20%
		All objects	3.77	+45.56%
	Sort by x coordinate	No distance	3.27	+26.25%
		Target-Distractor	**2.91**	**+12.36%**
		All objects	3.12	+20.46%
	Sort each coordinate	No distance	3.35	+29.34%
		Target-Distractor	3.33	+28.57%
		All objects	3.54	+36.68%
All objects	No sorting	No distance	3.57	+37.84%
		Target-Distractor	3.71	+43.24%
		All objects	4.03	+55.60%
	Sort by x coordinate	No distance	**3.01**	**+16.22%**
		Target-Distractor	**3.06**	**+18.15%**
		All objects	3.25	+25.48%
	Sort each coordinate	No distance	3.21	+23.94%
		Target-Distractor	3.44	+32.82%
		All objects	3.59	+38.61%
Previous eye gaze		-	0.27	-89.58%

Table 4.3: Median distance between the gaze and the predicted position from each model. Lower values in relative difference from the baseline represent more accurate predictions. The best predictions are emphasized.

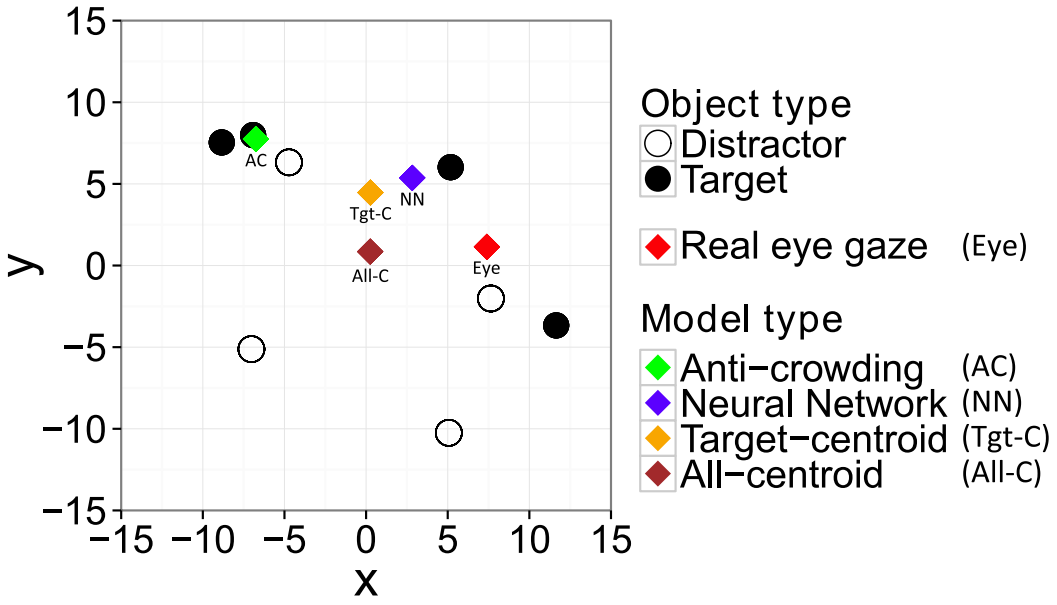


Figure 4.3: Real and predicted eye gaze in one particular sample. Filled-in points denote targets, empty points denote distractors, red diamonds denote eye gaze and other diamonds denote predicted position from several models. Reprinted with permission.

4.1.3 Discussion

We showed that machine learning for gaze prediction is possible in the MOT. We varied the amount of information available in the data and showed that we can get better predictions (in terms of Euclidean distance) than current models. Our results confirmed previous findings that eye gaze in MOT is not only driven only by targets but by distractors as well (Lukavský, 2013; Děchtěrenko & Lukavský, 2014). Because the best predictions were from the model that used only distance from targets to distractors instead of the position of distractors, some analytical models may be formed predicting eye gaze to the centroid of the targets; and thus, biased to the distractors. Such bias could be a variant of anti-crowding strategy (Děchtěrenko & Lukavský, 2014) or some other more complex variant.

Improvement of the models

Although the predictions from neural networks outperformed current strategies (by almost 10%), we need to study the predictions in detail to obtain a biologically plausible strategy. Although neural networks are just a black box, we

can analyze the samples in which the difference between analytical strategies and neural networks is biggest and modify current models to compensate for such differences. For example, there could be spatial configuration of the objects in which target-centroid would be a sufficient strategy. There could be other configurations (e.g. when targets and distractors are closer than critical spacing set by Bouma's distance; Bouma, 1973) that would lead to different strategies. A larger dataset for each experiment setting would allow us to train models on each of them separately and predict eye movements for each dataset separately. This could suggest which parameters of the experiments lead to different strategies. For example, evaluation of analytical strategies on eye gaze data from Experiments 3 and 4 showed that higher velocities of the objects decreased the performance of the centroid strategies in favor of strategies predicting eye gaze toward the center of the screen (Lux, 2014). When objects move at higher velocities, the tracking centroid of the virtual object may be more difficult and therefore saccades would be needed to keep up with the centroid. Because the scene is not processed during the saccades, object identity could be lost. Thus, it would be better to track the targets using only attention. Current models do not incorporate speed of the objects into them. Another aspect, which would lead to better performance of the models, would be to incorporate the fact that eye movements lag the scene content by approximately 110 ms (Lukavský & Děchtěrenko, 2016).

We trained the neural networks in the trials presented repeatedly and selected only those eye gaze samples that were closest to the median position for each time stamp. Because the variability of the scan patterns was still relatively high, the chance that future models will dramatically improve predictions is low. However, the coherence of scan patterns is still higher than scan patterns from free viewing of movies.

Models were evaluated by computing Euclidean distance between real eye gaze and predictions. Alternatively, we could evaluate scan patterns using a more complex metric such as correlation distance, which would capture the more complex properties of the scan patterns. For our purposes, per sample evaluation is appropriate, because our models worked with information for each sample individually.

Symmetry of the dataset

Neural networks trained using the dataset with sorted inputs had better prediction than unsorted data. For symmetrical input space, sorting the dataset is one plausible option for how to reduce the amount of data needed for the training. It is important to note that training using the sorted dataset would introduce some additional boundaries for the input space. These are hard to learn. For example, when we are training a machine learning algorithm to compute function $y = (x_1 + x_2)^2$ for the values $x_1, x_2 \in [-30, 30]$, we could train using the data set $[0, 30]$ for the cases $x_1 < x_2$, but we would get only half of the training cases near the boundary $x_1 = x_2$ in contrast with the original whole range. Interestingly, even when we sorted objects for each coordinate separately (and therefore the information about the proximity of objects was lost), we still got predictions better than those from the target-centroid strategy. Even without correspondence of the coordinates to the actual object, we could compute some sort of truncated or biased average for each coordinate separately. This could lead to strategies incorporating central bias of eye gaze (Tatler & Vincent, 2009).

Predicting eye gaze from targets versus predicting targets from eye gaze

Our results showed that information about the current state of the MOT task could be used for predicting eye gaze from target positions. This prediction relied only on the position of the targets and the distractors. Recently, Citoríik (2016) showed that neural networks could solve an inverse version of the problem. Can we predict which targets were tracked in the trial when the network has information about the position of the objects and eye gaze for the trial? This task was inspired by a similar study by Zelinsky and Todor (2010) in which support vector machines were trained to predict the category of the target during visual search tasks. In the case of MOT, predictions on target identity could be accumulated over all frames. In Citorik's work, he successfully trained neural networks and Hidden Markov models to predict target identity in MOT with almost perfect accuracy. Therefore, we conclude that scan patterns from dynamic tasks are a rich source of information and machine learning models could be trained using this data and

be used for prediction.

Use of predicting eye gaze for the scan pattern comparison

The importance of predicting eye gaze in MOT tasks helps us understand the variaiblity in the scan patterns in relationship to the tracking task. The eye gaze in MOT is influenced by three types of sources: top-down (participants plan their eye movement in order to track targets), bottom-up (the spatial configuration of the objects influences the fixated locations) and presence of internal noise. The development of new models for predicting eye gaze could help us explain the bottom-up influences on scan patterns. Therefore, we could represent scan patterns as time series in which we could express the eye gaze position as a combination of object coordinates and study the variability introduced by the natural noisiness of the eye movements. In addition, when scan patterns might be related to object positions, we could manipulate the trajectories of the objects in the task (similar to what we did in Experiments 3 and 4) and compare the differences in variability of the measured scans with the predicted ones.

Conclusion

In this thesis, we employed Multiple Object Tracking Task to study the properties of scan patterns with a prevalence of smooth pursuit eye movement. As such, scan patterns are basically time series. Our research has several conclusions for both computer science and vision science.

In the first chapter, we explored the properties of the Correlation distance metric in four simulations. This metric captures the similarity of groups of scan patterns in a natural way; it ranges from 0 to 1 and when the two groups of scan patterns share the same instances of some scan patterns, the metric scales appropriately (Simulation 2). This metric is highly correlated with Normalized Scanpath Saliency (Simulation 1), which we used in our previous experiments.

In the second part of the first chapter, we systematically modified scan patterns and computed similarity between the scan patterns and their modified versions. In particular, we selected three transformations: translation, rotation and scaling. Similar transformations have been used by other researchers for scan patterns in static tasks. We selected those particular transformations because they can easily be applied to the object trajectories in MOT as well.

In Simulation 3, we explored the robustness of Correlation distance with respect to the sample size used for our transformations. In Simulation 4, we explored the properties of metrics, when we applied either one or two transformations to a scan pattern at the same time. We compared similarities between the original and modified scan pattern using five metrics (Levenshtein, Fréchet, Correlation distance, Mean and Median).

Our results could be used by other researchers for comparison of results across experiments. In addition, the results from each behavioral experiment could be explained in terms of scan pattern variability. For example, when we observe the distance between scan patterns as .5 (when measured using Correlation distance), it would mean that the average distance between the scan patterns is 2°. For Simulations 1 and 2, we used artificial scan patterns, but Simulations 3 and 4 were done with scan patterns from an actual MOT experiment.

Our results could be used for different time series as well. The simulations could

be easily modified for different transformations and metrics. In the case of Simulations 1 and 2, the time series could be generated with different parameters to show resemblance to typical data in the given context. Alternatively, real time series could be used instead of generated ones. For Simulations 3 and 4, researchers could use metrics that are more common in the given context. We need to study how the metrics scale when we introduce artificial noise. Based on the typical variability of time series in each research context, researchers could interpret other studies and describe the variability in terms of transformations applied to the data.

In addition, since scan patterns can be treated as time series, researchers could employ additional metrics from different research areas. In this thesis, we showed an application of Fréchet distance that had not been used in this context before. It showed good properties for measuring the similarity of scan patterns. Similarly, other research fields that involve the comparison of time series could use metrics that are common for comparison of scan patterns. For example, Vyhlas (2016) applied the NSS metric for evaluating similarity of hand-written signatures in his thesis.

In the second chapter, we studied methods for testing statistical differences between two groups of scan patterns. For cases where we are interested in the overall difference between groups, we could use classic statistical methods. We showed an example of this design in Experiment 1. For cases where we want to test when variability between groups is different from variability within groups, we have to use a different approach. In particular, we developed two methods for testing the statistical significance of comparisons between two groups of scan patterns. In the Simulated Experiment 2, we tested the properties of our two methods and the method that is currently used to test the differences between groups. In Experiments 3 and 4, we showed the application of our novel method (we used the best-performing method based on the results of Simulation 1). Our approach could be used to test differences between groups of time series in general.

All of the Experiments 1, 3, and 4 also delivered results that are beneficial for the vision science community. In Experiment 1, we tested the effect of wearing glasses

on the coherence of the group of scan patterns. In Experiments 3 and 4, we tested the possibility of flipping object trajectories in MOT tasks. This addresses the theoretical question about the symmetry of scan patterns and practical questions on the possibility of masking repetition. Although our two methods performed similar to the original one, they work with the coherence of the whole group of scan patterns instead of averaged distance between all pairs of scan patterns.

In the final chapter, we predicted scan patterns in MOT using neural networks. We varied the feature vector that we used for the training and we showed that distance between targets and distractors is an important feature that influences the quality of predictions. Our results are in concordance with our previous results. However, in this scenario, we used a data-driven approach with no a priori theory. Scan patterns predicted by the neural networks outperformed the current analytical strategies for predicting eye gaze in MOT. Predicting scan patterns in MOT is an important goal. If we could explain eye gaze with respect to the position of objects, we could express scan patterns as time series with objects as independent variables. By modifying trajectories through adding systematic noise, we could study the differences in scan pattern variability in comparison to predicted variability.

Ultimately, we showed a non-traditional use of a MOT task for various research questions. MOT is an ideal task for studying the properties of scan pattern comparison metrics. We could present trials repeatedly without noticing and study intra-subject variability. This is not possible when using movies as stimuli. In MOT, we could systematically modify the object trajectories and see how this transformation projects itself in scan patterns.

We hope that our work encourages other researchers to study scan patterns in dynamic tasks. Dynamic stimuli are currently underused in eye tracking research due to the complexity of analysis and lack of methods for testing significance between groups of scan patterns. We believe that our research will provide an additional piece of the puzzle as concerns questions of the quality of eye data.

List of Author's Publications

A) Publications whose results were used in this thesis:

1. **Děchtěrenko, F.**, Lukavský, J., & Holmqvist, K. (2017). Flipping the stimulus: Effects on scanpath coherence? *Behavior Research Methods*, 49(1), 382–393. IF 2015: 3.048

2. Lukavský, J., & **Děchtěrenko, F.** (2016). Gaze position lagging behind scene content in multiple object tracking: Evidence from forward and backward presentations. *Attention, Perception, & Psychophysics*, 78(8), 2456–2468. IF 2015: 1.782

3. **Děchtěrenko, F.**, & Lukavský, J. (2014). Models of eye movements in multiple object tracking with many objects. In *5th European Workshop on Visual Information Processing (EUVIP 2014)* (pp. 1–6). IEEE.

4. **Děchtěrenko, F.**, & Lukavský, J. (2016). Predicting eye movements in multiple object tracking using neural networks. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (pp. 271–274).

B) Supervised theses which were referenced in this thesis:

1. Citorík, J. (2016). *Predicting targets in Multiple Object Tracking task.* (Master's thesis, Charles University, Prague).

2. Kocián, M. (2014). *Metriky pro porovnávání očních pohybů.* (Bachelor's thesis, Charles University, Prague).

3. Lux, E. (2014). *Bayesian models of eye movements.* (Master's thesis, Charles University, Prague).

4. Vyhlas, P. (2016). *Porovnávání podpisů pomocí metrik pro oční pohyby.* (Bachelor's thesis, Charles University, Prague).

C) Other Author's publications

1. Bojar, O., **Děchtěrenko, F.**, & Zelenina, M. (2016). Eye-Tracking Recordings from a Pilot Study of WMT-style MT Outputs Ranking.

2. Černý, V., & **Děchtěrenko, F.** (2015). Rogue-Like Games as a Playground for Artificial Intelligence – Evolutionary Approach. In *Entertainment Computing - ICEC 2015* (pp. 261–271). Springer International Publishing.

3. Beneš, J., Kelly, T., **Děchtěrenko, F.**, Křivánek, J., Müller, P., On Realism of Architectural Procedural Models. *Computer Graphics Forum* (to appear). IF 2015: 1.476

4. Brom, C., Buchtová, M., Šisler, V., **Děchtěrenko, F.**, Palme, R., & Glenk, L. M. (2014). Flow, social interaction anxiety and salivary cortisol responses in serious games: A quasi-experimental study. *Computers & Education*, 79, 69–100. IF 2015: 2.881

5. Brom, C., Bromová, E., **Děchtěrenko, F.**, Buchtová, M., & Pergel, M. (2014). Personalized messages in a brewery educational simulation: Is the personalization principle less robust than previously thought? *Computers & Education*, 72, 339–366. IF 2015: 2.881

6. Brom, C., Hannemann, T., Stárková, T., Bromová, E., & **Děchtěrenko, F.** (in print). The role of cultural background in the personalization principle: Five experiments with Czech learners. *Computers & Education*. IF 2015: 2.881

7. Brom, C., & **Děchtěrenko, F.** (2015). Mathematical Self-Efficacy as a Determinant of Successful Learning of Mental Models From Computerized Materials. In *ECGBL2015-9th European Conference on Games Based Learning: ECGBL2015* (pp. 89–97).

References

Abed, F. (1991). Cultural influences on visual scanning patterns. *Journal of Cross-Cultural Psychology*, *22*(4), 525–534. doi: 10.1177/0022022191224006

Agtzidis, I., Startsev, M., & Dorr, M. (2016). Smooth pursuit detection based on multiple observers. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '16* (pp. 303–306). New York, New York, USA: ACM Press. doi: 10.1145/2857491.2857521

Al-Rahayfeh, A., & Faezipour, M. (2013). Eye tracking and head movement detection: A state-of-art survey. *IEEE Journal of Translational Engineering in Health and Medicine*, *1*, 2100212–2100212. doi: 10.1109/JTEHM.2013.2289879

Alt, H., & Godau, M. (1995). Computing the Frechét distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, *05*(01n02), 75–91. doi: 10.1142/S0218195995000064

Alvarez, G. A. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643. doi: 10.1111/j.1467-9280.2005.01587.x

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13). doi: 10.1167/7.13.14

Anderson, N. C., Bischof, W. F., Laidlaw, K. E. W., Risko, E. F., & Kingstone, A. (2013). Recurrence quantification analysis of eye movements. *Behavior Research Methods*, *45*(3), 842–856. doi: 10.3758/s13428-012-0299-5

Anderson, N. C., Laidlaw, K. E. W., Bischof, W. F., & Kingstone, A. (2012). Recurrence quantification analysis of scan patterns. *Journal of Vision*, *12*(9), 544–544. doi: 10.1167/12.9.544

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (n.d.). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*. doi: 10.3758/s13428-016-0738-9

Bach, M. (2007). The Freiburg Visual Acuity Test-Variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *245*(7), 965–971. doi: 10.1007/s00417-006-0474-4

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition*, *10*(8), 949–963. doi: 10.1080/13506280344000158

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2007). *Neuroscience: Exploring the brain* (3rd ed.). Lippincott Williams & Wilkins.

Berndt, D., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *Workshop on Knowledge Knowledge Discovery in Databases*, *398*, 359–370.

Blignaut, P., & Wium, D. (2014). Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior Research Methods*, *46*(1), 67–80. doi: 10.3758/s13428-013-0343-0

Bonnen, K., Burge, J., Yates, J., Pillow, J., & Cormack, L. K. (2015). Continuous psychophysics: Target-tracking to measure visual sensitivity. *Journal of Vision*, *15*(3), 14–14. doi: 10.1167/15.3.14

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research*, *13*(4), 767–782. doi: 10.1016/0042-6989(73)90041-2

Bowers, D., & Heilman, K. M. (1980). Pseudoneglect: Effects of hemispace on a tactile line bisection task. *Neuropsychologia*, *18*(4-5), 491–498. doi: 10.1016/0028-3932(80)90151-7

Bradley, C., Abrams, J., & Geisler, W. S. (2014). Retina-V1 model of detectability across the visual field. *Journal of Vision*, *14*(12), 22–22. doi: 10.1167/14.12.22

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. doi: 10.1073/pnas.0803390105

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. doi: 10.1163/156856897X00357

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, *9*(1), 27–38. doi: 10.1162/jocn.1997.9.1.27

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*(6), 719–722. doi: 10.1016/0042-6989(75)90290-4

Burr, D. C., Morrone, M. C., & Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature*, *371*(6497), 511–513. doi: 10.1038/371511a0

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*(7), 349–354. doi: 10.1016/j.tics.2005.05.009

Citorík, J. (2016). *Predicting targets in Multiple Object Tracking task*. Master's thesis, Charles University.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Colas, F., Flacher, F., Tanner, T., Bessière, P., & Girard, B. (2009). Bayesian models of eye movement selection with retinotopic maps. *Biological Cybernetics*, *100*(3), 203–214. doi: 10.1007/s00422-009-0292-y

Corballis, P. (2002). Hemispheric asymmetries for simple visual judgments in the split brain. *Neuropsychologia*, *40*(4), 401–410. doi: 10.1016/S0028-3932(01)00100-2

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 613–617. doi: 10.3758/BF03195489

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), S0140525X01003922. doi: 10.1017/S0140525X01003922

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, *42*(3), 692–700. doi: 10.3758/BRM.42.3.692

De Valois, R. L., & De Valois, K. K. (1980). Spatial Vision. *Annual Review of Psychology*, *31*(1), 309–341. doi: 10.1146/annurev.ps.31.020180.001521

Děchtěrenko, F. (2012). *Modelling eye movements during Multiple Object Tracking*. Master's thesis, Charles University in Prague.

Děchtěrenko, F., & Lukavský, J. (2014). Models of eye movements in multiple object tracking with many objects. In *5th European Workshop on Visual Information Processing (EUVIP 2014)* (pp. 1–6). IEEE. doi: 10.1109/EUVIP.2014.7018375

Děchtěrenko, F., & Lukavský, J. (2016). Predicting eye movements in multiple object tracking using neural networks. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '16* (pp. 271–274).

Děchtěrenko, F., Lukavský, J., & Holmqvist, K. (2017). Flipping the stimulus: Effects on scanpath coherence? *Behavior Research Methods*, *49*(1), 382–393. doi: 10.3758/s13428-016-0708-2

Dempere-Marco, L., Hu, X.-P., Ellis, S., Hansell, D., & Yang, G.-Z. (2006). Analysis of visual search patterns with EMD metric in normalized anatomical space. *IEEE Transactions on Medical Imaging*, *25*(8), 1011–1021. doi: 10.1109/TMI.2006.875427

Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behavior Research Methods*, *44*(4), 1079–1100. doi: 10.3758/s13428-012-0212-2

Dickinson, C. A., & Intraub, H. (2009). Spatial asymmetries in viewing and remembering scenes: Consequences of an attentional bias? *Attention, Perception, & Psychophysics*, *71*(6), 1251–1262. doi: 10.3758/APP.71.6.1251

Dodge, R. (1900). Visual perception during eye movement. *Psychological Review*, *7*(5), 454–465. doi: 10.1037/h0067215

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28. doi: 10.1167/10.10.28

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373). Springer Science & Business Media.

Evdokimidis, I., Smyrnis, N., Constantinidis, T., Stefanis, N., Avramopoulos, D., Paximadis, C., ... Stefanis, C. (2002). The antisaccade task in a sample of 2,006 young men: I. Normal population characteristics. *Experimental Brain Research*, *147*(1), 45–52. doi: 10.1007/s00221-002-1208-4

Fehd, H. M. (2009). *Eye movement strategies during attentional tracking.* Unpublished doctoral dissertation, Vanderbilt University.

Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: where do participants look? *Cognition*, *108*(1), 201–9. doi: 10.1016/j.cognition.2007.11.008

Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, *10*(4), 1–13. doi: 10.1167/10.4.19

Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception & Psychophysics*, *69*(4), 567–577. doi: 10.3758/BF03193914

Feusner, M., & Lukoff, B. (2008). Testing for statistically significant differences between groups of scan patterns. In *Proceedings of the Fifth Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '08* (p. 43). New York, New York, USA: ACM Press. doi: 10.1145/1344471.1344481

Finke, R. A., & Kosslyn, S. M. (1980). Mental imagery acuity in the peripheral visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 126–139. doi: 10.1037/0096-1523.6.1.126

Fischer, B., & Ramsperger, E. (1984). Human express saccades: extremely short reaction times of goal directed eye movements. *Experimental Brain Research*, *57*(1). doi: 10.1007/BF00231145

Foulsham, T., Gray, A., Nasiopoulos, E., & Kingstone, A. (2013). Leftward biases in picture scanning and line bisection: A gaze-contingent window study. *Vision Research*, *78*, 14–25. doi: 10.1016/j.visres.2012.12.001

Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: effects of image type and image

features. *Vision Research*, *50*(8), 779–95. doi: 10.1016/j.visres.2010.01.019

Foulsham, T., & Kingstone, A. (2013). Fixation-dependent memory for natural scenes: An experimental test of scanpath theory. *Journal of Experimental Psychology: General*, *142*(1), 41–56. doi: 10.1037/a0028227

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: patterns in saccade direction vary with picture orientation. *Vision Research*, *48*(17), 1777–90. doi: 10.1016/j.visres.2008.05.018

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science*, *21*(7), 920–925.

Franconeri, S. L., Lin, J. Y., Pylyshyn, Z. W., Fisher, B., & Enns, J. T. (2008). Evidence against a speed limit in multiple-object tracking. *Psychonomic Bulletin & Review*, *15*(4), 802–808.

Fréchet, M. M. (1906). Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo*, *22*(1), 1–72. doi: 10.1007/BF03018603

Freeman, R. D. (1980). Visual acuity is better for letters in rows than in columns. *Nature*, *286*(5768), 62–64. doi: 10.1038/286062a0

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: John Wiley and Sons.

Greene, H. H., Brown, J. M., & Dauphin, B. (2014). When do you look where you look? A visual field asymmetry. *Vision Research*, *102*, 33–40. doi: 10.1016/j.visres.2014.07.012

Hagenbeek, R. E., & Van Strien, J. W. (2002). Left-right and upper-lower visual field asymmetries for face matching, letter naming, and lexical decision. *Brain and Cognition*, *49*(1), 34–44. doi: 10.1006/brcg.2001.1481

Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194. doi: 10.1016/j.tics.2005.02.009

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 6. doi: 10.1167/3.1.6

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking : A comprehensive guide to methods and*

measures. New York: OUP Oxford.

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality. In *Proceedings of the Seventh Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '12* (Vol. 1, p. 45). New York, New York, USA: ACM Press. doi: 10.1145/2168556.2168563

Howe, P. D. L., & Holcombe, A. O. (2012). Motion information is sometimes used as an aid to the visual tracking of objects. *Journal of Vision*, *12*, 1–10. doi: 10.1167/12.13.10.Introduction

Hubel, D. H., & Wiesel, T. N. (1974). Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor. *The Journal of Comparative Neurology*, *158*(3), 295–305. doi: 10.1002/cne.901580305

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*(3), 171–216. doi: 10.1006/cogp.2001.0755

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. doi: 10.1109/34.730558

Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. In *Proceedings of the Sixth Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '10* (p. 211). New York, New York, USA: ACM Press. doi: 10.1145/1743666.1743718

Jewell, G., & McCourt, M. E. (2000). Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia*, *38*(1), 93–110. doi: 10.1016/S0028-3932(99)00045-7

Jiang, M., Xu, Y., & Zhu, B. (2008). Protein structure–structure alignment with discrete Fréchet distance. *Journal of Bioinformatics and Computational Biology*, *06*(01), 51–64. doi: 10.1142/S0219720008003278

Jost, T., Ouerhani, N., Wartburg, R. V., Müri, R., & Hügli, H. (2005). Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding*, *100*(1-2), 107–123. doi: 10.1016/j.cviu.2004.10.009

Ke, S. R., Lam, J., Pai, D. K., & Spering, M. (2013). Directional asymmetries

in human smooth pursuit eye movements. *Investigative Opthalmology & Visual Science*, *54*(6), 4409. doi: 10.1167/iovs.12-11369

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology*, *52*(4), 346–68. doi: 10.1016/j.cogpsych.2005.12.001

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox. *Perception*, *36*(ECVP Abstract Supplement), 14–14. doi: 10.1068/v070821

Kocián, M. (2014). *Metriky pro porovnávání očních pohybů*. Bachelor's thesis, Charles University.

Komogortsev, O. V., & Karpov, A. (2013). Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior Research Methods*, *45*(1), 203–215. doi: 10.3758/s13428-012-0234-9

Komogortsev, O. V., & Khan, J. I. (2008). Eye movement prediction by Kalman filter with integrated linear horizontal oculomotor plant mechanical model. In *Proceedings of the Fifth Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '08* (p. 229). New York, New York, USA: ACM Press. doi: 10.1145/1344471.1344525

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*(6483), 742–744. doi: 10.1038/369742a0

Landry, S. J., Sheridan, T. B., & Yufik, Y. M. (2001). A methodology for studying cognitive groupings in a target-tracking task. *IEEE Transactions on Intelligent Transportation Systems*, *2*(2), 92–100. doi: 10.1109/6979.928720

Larsson, L., Nyström, M., Andersson, R., & Stridh, M. (2015). Detection of fixations and smooth pursuit movements in high-speed eye-tracking data. *Biomedical Signal Processing and Control*, *18*, 145–152. doi: 10.1016/j.bspc.2014.12.008

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research Methods*, *45*(1), 251–66. doi: 10.3758/s13428-012-0226-9

Le Meur, O., Le Callet, P., Barba, D., & Thoreau, D. (2006). A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(5), 802–817. doi: 10.1109/TPAMI.2006.86

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654. doi: 10.1016/j.visres.2007.12.009

Levi, D. M., Song, S., & Pelli, D. G. (2007). Amblyopic reading is crowded. *Journal of Vision*, *7*(2), 21.1–17. doi: 10.1167/7.2.21

Levine, M. W., & McAnany, J. J. (2005). The relative capabilities of the upper and lower visual hemifields. *Vision Research*, *45*(21), 2820–2830. doi: 10.1016/j.visres.2005.04.001

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., & Enns, J. T. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 235–247. doi: 10.1037/0096-1523.31.2.235

Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLOS ONE*, *10*(11), e0142474. doi: 10.1371/journal.pone.0142474

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. doi: 10.1038/36846

Lukavský, J. (2013). Eye movements in repeated multiple object tracking. *Journal of Vision*, *13*(7), 9–9. doi: 10.1167/13.7.9

Lukavský, J., & Děchtěrenko, F. (2016). Gaze position lagging behind scene content in multiple object tracking: Evidence from forward and backward presentations. *Attention, Perception, & Psychophysics*, *78*(8), 2456–2468. doi: 10.3758/s13414-016-1178-4

Lux, E. (2014). *Bayesian models of eye movements.* Master's thesis, Charles

University.

Mannan, S., Ruddock, K., & Wooding, D. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision*, *9*(3), 363–386. doi: 10.1163/156856895X00052

Mannan, S., Ruddock, K., & Wooding, D. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, *10*(3), 165–188. doi: 10.1163/156856896X00123

Mannan, S., Ruddock, K., & Wooding, D. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision*, *11*(2), 157–178. doi: 10.1163/156856897X00177

Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*. Chichester: Wiley.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, *11*(2), 431–441. doi: 10.1137/0111030

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*(3), 229–240. doi: 10.1038/nrn1348

Marwan, N., & Kurths, J. (2002). Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A*, *302*(5-6), 299–307. doi: 10.1016/S0375-9601(02)01170-2

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. doi: 10.3758/s13428-011-0168-7

Meyerhoff, H. S., Papenmeier, F., Jahn, G., & Huff, M. (2015). Distractor locations influence Multiple Object Tracking beyond interobject spacing. *Experimental Psychology*, *62*(3), 170–180. doi: 10.1027/1618-3169/a000283

Molitor, R. J., Ko, P. C., & Ally, B. A. (2017). Eye Movements in Alzheimer's Disease. *Journal of Alzheimer's Disease*, *44*(1), 1–12. doi: 10.3233/JAD-141173

Monti, M. (2011). Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach. *Frontiers in Human Neuroscience*, *5*(609), 28. doi:

10.3389/fnhum.2011.00028

Najemnik, J., & Geisler, W. S. (2009). Simple summation rule for optimal fixation selection in visual search. *Vision Research*, *49*(10), 1286–1294. doi: 10.1016/j.visres.2008.12.005

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, *48*(3), 443–453. doi: 10.1016/0022-2836(70)90057-4

Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science*, *171*(3968), 308–311. doi: 10.1126/science.171.3968.308

Nuthmann, A., & Matthias, E. (2014). Time course of pseudoneglect in scene viewing. *Cortex*, *52*, 113–119. doi: 10.1016/j.cortex.2013.11.007

Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, *45*(1), 272–288. doi: 10.3758/s13428-012-0247-4

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, *42*(1), 188–204. doi: 10.3758/BRM.42.1.188

Ogawa, H., Watanabe, K., & Yagi, A. (2009). Contextual cueing in multiple object tracking. *Visual Cognition*, *17*(8), 1244–1258. doi: 10.1080/13506280802457176

Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition*, *11*(5), 631–671. doi: 10.1080/13506280344000473

Ossandon, J. P., Onat, S., & Konig, P. (2014). Spatial biases in viewing behavior. *Journal of Vision*, *14*(2), 20–20. doi: 10.1167/14.2.20

Papenmeier, F., & Huff, M. (2010). DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animations and movies. *Behavior Research Methods*, *42*(1), 179–187. doi: 10.3758/BRM.42.1.179

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal*

of Neuroscience Methods, *162*(1-2), 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. doi: 10.1163/156856897X00366

Peters, R. J., & Itti, L. (2008). Applying computational tools to predict gaze direction in interactive visual environments. *ACM Transactions on Applied Perception*, *5*(2), 1–19. doi: 10.1145/1279920.1279923

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*(18), 2397–2416. doi: 10.1016/j.visres.2005.03.019

Petrov, Y., & Meleshkevich, O. (2011). Asymmetries and idiosyncratic hot spots in crowding. *Vision Research*, *51*(10), 1117–1123. doi: 10.1016/j.visres.2011.03.001

Pitzalis, S., & Di Russo, F. (2001). Spatial anisotropy of saccadic latency in normal subjects and brain-damaged patients. *Cortex*, *37*(4), 475–492. doi: 10.1016/S0010-9452(08)70588-4

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, *109*(2), 160–174. doi: 10.1037/0096-3445.109.2.160

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*(1), 65–97. doi: 10.1016/0010-0277(89)90014-0

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, *11*(7), 801–822. doi: 10.1080/13506280344000518

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197. doi: 10.1163/156856888X00122

R Core Team. (2016). *R: A language and environment for statistical computing.* Vienna, Austria. Retrieved from `http://www.r-project.org/`

Rajashekar, U., Cormack, L. K., & Bovik, A. C. (2004). Point of gaze analysis

reveals visual search strategies. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human Vision and Electronic Imaging IX* (Vol. 5292, pp. 296–306). doi: 10.1117/12.537118

Rajashekar, U., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2008). GAFFE: a gaze-attentive fixation finding engine. *IEEE transactions on Image Processing*, *17*(4), 564–73. doi: 10.1109/TIP.2008.917218

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi: 10.1037/0033-2909.124.3.372

Riche, N., Duvinage, M., Mancas, M., Gosselin, B., & Dutoit, T. (2013). Saliency and human fixations: State-of-the-art and study of comparison metrics. *Proceedings of the IEEE International Conference on Computer Vision*, 1153–1160. doi: 10.1109/ICCV.2013.147

Robinson, D. A. (1963). A method of measuring eye movemnent using a scieral search coil in a magnetic field. *IEEE Transactions on Bio-medical Electronics*, *10*(4), 137–145. doi: 10.1109/TBMEL.1963.4322822

Rommelse, N. N., Van der Stigchel, S., & Sergeant, J. A. (2008). A review on eye movement studies in childhood and adolescent psychiatry. *Brain and Cognition*, *68*(3), 391–414. doi: 10.1016/j.bandc.2008.08.025

Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision*, *3*(1), 2–2. doi: 10.1167/3.1.2

Salvucci, D. D., & Anderson, J. R. (1998). Tracing eye movement protocols with cognitive process models. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, 923–928.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Second Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '00* (pp. 71–78). New York, New York, USA: ACM Press. doi: 10.1145/355017.355028

Santini, T., Fuhl, W., Kübler, T., & Kasneci, E. (2016). Bayesian identification of fixations, saccades, and smooth pursuits. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '16* (pp. 163–170). New York, New York, USA: ACM Press. doi:

10.1145/2857491.2857512

Schleicher, R., Galley, N., Briest, S., & Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*, *51*(7), 982–1010. doi: 10.1080/00140130701817062

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*(1-2), 159–177. doi: 10.1016/S0010-0277(00)00157-8

Schutz, A. C., Braun, D. I., Kerzel, D., & Gegenfurtner, K. R. (2008). Improved visual sensitivity during smooth pursuit eye movements. *Nature Neuroscience*, *11*(10), 1211–1216. doi: 10.1038/nn.2194

Scimeca, J. M., & Franconeri, S. L. (2015). Selecting and tracking multiple objects. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(2), 109–118. doi: 10.1002/wcs.1328

Sheliga, B., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research*, *98*(3). doi: 10.1007/BF00233988

Shim, W. M., Alvarez, G. A., & Jiang, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin & Review*, *15*(2), 390–397. doi: 10.3758/PBR.15.2.390

Smit, A. C., & Van Gisbergen, J. A. M. (1990). An analysis of curvature in fast and slow human saccades. *Experimental Brain Research*, *81*(2), 335–345. doi: 10.1007/BF00228124

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*(8), 16–16. doi: 10.1167/13.8.16

Sriraghavendra, E., Karthik, K., & Bhattacharyya, C. (2007). Fréchet distance based approach for searching online handwritten documents. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)* (Vol. 1, pp. 461–465). IEEE. doi: 10.1109/ICDAR.2007.4378752

St.Clair, R. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision*, *10*(4), 1–13. doi: 10.1167/10.4.18

Suchow, J. W., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms

of the debate on the format and structure of visual memory. *Attention, Perception, & Psychophysics*. doi: 10.3758/s13414-014-0690-7

Tafaj, E., Kasneci, G., Rosenstiel, W., & Bogdan, M. (2012). Bayesian online clustering of eye movement data. In *Proceedings of the Seventh Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '12* (p. 285). New York, New York, USA: ACM Press. doi: 10.1145/2168556.2168617

Tang, H., Topczewski, J. J., Topczewski, A. M., & Pienta, N. J. (2012). Permutation test for groups of scanpaths using normalized Levenshtein distances and application in NMR questions. In *Proceedings of the Seventh Biennial ACM Symposium on Eye Tracking Research & Applications - ETRA '12* (p. 169). New York, New York, USA: ACM Press. doi: 10.1145/2168556.2168584

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, *45*(5), 643–59. doi: 10.1016/j.visres.2004.09.017

Tatler, B. W., & Hutton, S. B. (2007). Trial by trial effects in the antisaccade task. *Experimental Brain Research*, *179*(3), 387–396. doi: 10.1007/s00221-006-0799-6

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*(6-7), 1029–1054. doi: 10.1080/13506280902764539

Toet, A. (2011). Computational versus psychophysical bottom-up image saliency: A comparative evaluation study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(11), 2131–2146. doi: 10.1109/TPAMI.2011.53

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, *32*(7), 1349–1357. doi: 10.1016/0042-6989(92)90227-A

Tombu, M., & Seiffert, A. E. (2011). Tracking planets and moons: mechanisms of object tracking revealed with a new paradigm. *Attention, Perception, & Psychophysics*, *73*(3), 738–50. doi: 10.3758/s13414-010-0060-z

Treisman, A., & Zhang, W. (2006). Location and binding in visual working mem-

ory. *Memory & Cognition*, *34*(8), 1704–1719. doi: 10.3758/BF03195932

Vul, E., Frank, M. C., Tenenbaum, J. B., & Alvarez, G. A. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. In *Advances in Neural Information Processing Systems* (pp. 1955–1963).

Vyhlas, P. (2016). *Porovnávání podpisů pomocí metrik pro oční pohyby*. Bachelor's thesis, Charles University.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, *24*(3), 295–340. doi: 10.1016/0010-0285(92)90010-Y

Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press.

Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, *16*(5), 553–566. doi: 10.1080/13506280802000752

Zelinsky, G. J., & Todor, A. (2010). The role of "rescue saccades" in tracking objects through occlusions. *Journal of Vision*, *10*(14), 1–13. doi: 10.1167/10.14.29

Zhang, Y., & Hornof, A. (2011). Mode-of-disparities error correction of eye-tracking data. *Behavior Research Methods*, *43*(3), 834–842. doi: 10.3758/s13428-011-0073-0

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Full name
AOI	Areas of interests
CD	Correlation Distance
DVA	Degrees of visual angle
MOT	Multiple Object Tracking
NSS	Normalized Scanpath Saliency
ROC	Receiver Operator Characteristic

Attachments

Appendix A – Transformation tables

In this section, we show tables that relate the different distortions of scan patterns to each other. These tables could be used for comparison of the results of different studies. Tables A.1, A.2, and A.3 shows average values for five metrics (CD, Fréchet, Levenshtein, Mea, and Median) for each of the three transformations (translation, rotation, and scaling). Tables A.4, A.5, and A.6 relate the application of one transformation to another in the values for each metric. See detailed description in Section 2.10.2.

Translation	CD	Fréchet	Levenshtein	Mean	Median
0.25	0.01	0.25	0.10	0.25	0.25
0.50	0.04	0.50	0.22	0.50	0.50
0.75	0.09	0.75	0.30	0.75	0.75
1.00	0.16	1.00	0.40	1.00	1.00
2.00	0.50	2.00	0.77	2.00	2.00
3.00	0.80	3.00	0.96	3.00	3.00

Table A.1: Mean values of each metric for the translation.

Angle	CD	Fréchet	Levenshtein	Mean	Median
5.00	0.04	0.77	0.18	0.44	0.41
10.00	0.15	1.52	0.39	0.88	0.82
15.00	0.28	2.26	0.54	1.32	1.23
20.00	0.41	2.99	0.64	1.76	1.64

Table A.2: Mean values of each metric for the rotation.

Scale	CD	Fréchet	Levenshtein	Mean	Median
0.50	0.59	4.32	0.78	2.53	2.36
0.75	0.27	2.25	0.56	1.26	1.18
0.90	0.06	0.91	0.26	0.51	0.47

Table A.3: Mean values of each metric for the scaling.

	Rotation				
Translation	CD	Fréchet	Levenshtein	Mean	Median
0.25 DVA	2.36	1.59	2.73	2.83	3.04
0.50 DVA	4.88	3.22	6.00	5.66	6.08
0.75 DVA	7.62	4.85	7.97	8.50	9.12
1.00 DVA	10.33	6.50	10.36	11.34	12.17
2.00 DVA	24.32	13.22	28.70	22.80	24.48
	Scaling				
Translation	CD	Fréchet	Levenshtein	Mean	Median
0.25 DVA	0.98	0.97	0.96	0.95	0.95
0.50 DVA	0.92	0.95	0.92	0.90	0.89
0.75 DVA	0.87	0.92	0.88	0.85	0.84
1.00 DVA	0.82	0.89	0.84	0.80	0.79
2.00 DVA	0.58	0.78	0.51	0.60	0.58

Table A.4: Corresponding rotation (degrees) and scaling (scaling factor) transformation for each step of translation operation.

	Translation				
Rotation	CD	Fréchet	Levenshtein	Mean	Median
5°	0.51	0.77	0.43	0.44	0.41
10°	0.97	1.52	0.97	0.88	0.82
15°	1.38	2.26	1.43	1.32	1.23
20°	1.73	2.99	1.65	1.76	1.64
	Scaling				
Rotation	CD	Fréchet	Levenshtein	Mean	Median
5°	0.92	0.92	0.93	0.91	0.91
10°	0.83	0.83	0.85	0.83	0.83
15°	0.74	0.75	0.76	0.74	0.74
20°	0.65	0.66	0.67	0.65	0.65

Table A.5: Corresponding translation (DVA) and scaling transformation (scaling factor) for each step of rotation operation.

	Translation				
Scaling	CD	Fréchet	Levenshtein	Mean	Median
0.50	2.25		2.04	2.53	2.36
0.75	1.33	2.25	1.47	1.26	1.18
0.90	0.57	0.91	0.60	0.51	0.47
	Rotation				
Scaling	CD	Fréchet	Levenshtein	Mean	Median
0.50	29.02	29.97	29.47	28.96	28.96
0.75	14.33	14.93	15.64	14.36	14.36
0.90	5.71	5.91	7.05	5.73	5.73

Table A.6: Corresponding translation (DVA) and rotation transformation (degrees) for each step of scaling operation.

Appendix B – Code description

As most of the analysis was performed in language R, we developed the package *scanpatterns*. This package contains scripts that represent scan patterns as a S3 object. Inside the package, there are also comparison methods used in this thesis. The package can be downloaded from github.com (`https://github.com/fidadoma/scanpatterns`) and it can be installed in R using command devtools:: install_github ("fidadoma/scanpatterns").

We stored the scripts used for analysis as well as the data on the Open Science Framework on the url address `https://osf.io/ek5by/`.